AD-A056 853 STANFORD UNIV CALIF STANFORD ELECTRONICS LABS F/G 17/2
SPECTRAL DISTORTION MEASURES FOR SPEECH COMPRESSION.(U)
APR 78 Y MATSUYAMA, A BUZO, R M GRAY F44620-73-C-0065
UNCLASSIFIED SU-SEL-78-015 NL

1 OF 1
ADA
056853

END
DATE
FILMED
9-78
DDC

INFORMATION SYSTEMS LABORATORY

STANFORD ELECTRONICS LABORATORIES
DEPARTMENT OF ELECTRICAL ENGINEERING
STANFORD UNIVERSITY · STANFORD, CA 94305

LEVEL II

SEL-78-015

12

SPECTRAL DISTORTION MEASURES FOR SPEECH COMPRESSION

by

Y. Matsuyama, A. Buzo, and R.M. Gray

April 1978

Technical Report No. 6504-3

Prepared under

U.S. Air Force Office of Scientific Research
Contract F44620-73-C-0065 and by the
Joint Services Electronics Program
(U.S. Army, U.S. Navy, and U.S. Air Force)
under Contract N00014-75-C-0601

14 SU-SEL-78-015, TR-6504-3

6 SPECTRAL DISTORTION MEASURES FOR SPEECH COMPRESSION,

by

10 Y. Matsuyama, A. Buzo [redacted] R. M. Gray

11 Apr[redacted] [redacted]978   12 58p.

9 Technical rept. 1 May 73-30 Apr 78,

Technical Report No. 6504-3

Prepared under

U.S. Air Force Office of Scientific Research
Contract F44620-73-C-0065 and by the
Joint Services Electronics Program
(U.S. Army, U.S. Navy, and U.S. Air Force)
under Contract N00014-75-C-0601

15 F44620-73-C-0065,
N00014-75-C-0601

Information Systems Laboratory
Stanford Electronics Laboratories
Stanford University   Stanford, California

332 400

SPECTRAL DISTORTION MEASURES FOR SPEECH COMPRESSION

by

Y. Matsuyama, A. Buzo, and R.M. Gray‡

ABSTRACT

In recent years several measures of distortion between speech waveforms have been proposed as substitutes for the traditional but subjectively inadequate mean-squared error. All of these measures involve some form of distortion measure between the second order properties of the speech processes producing the waveforms instead of an average of the waveform error power. In particular, they depend on the power spectral densities or linear models of the speech process. In this report the properties and interrelations of several such measures are developed. In particular, the relative strengths or equivalences of the various implications and applications of these measures to prediction, detection, and coding are summarized.

---

‡The authors are with the Information Systems Laboratory, Stanford University, Stanford, CA 94305. This research was supported by the Air Force Office of Scientific Research under Contract F44620-73-C-0065 and by the Joint Services Electronics Program at Stanford under contract N00014-75-C-0601.

1. INTRODUCTION

Any communications system for human speech has a natural fidelity criterion -- the subjective fidelity for a given customer, that is, whether or not the final reconstructed speech sounds "good" or "bad." For several reasons, however, it is desirable to have a mathematical fidelity criterion -- a formula for computing a number from two speech waveforms that measures the "distortion" or "badness of approximation" between them. Such a mathematical criterion provides an absolute yardstick independent of individual listeners' differences of taste and may allow the theoretical analysis of such systems, e.g., the application of communications theory to develop "optimal" performance bounds with which to compare actual system performance. In addition, a distortion measure can play a crucial role in the actual operation of the communication systems or, for example, in coding or generalized quantization systems where one selects a reproduction symbol from an allowed set by choosing the one having minimum distortion from the given input symbol.

To be useful, any such criterion must possess to some degree the following attributes: (1) It should be subjectively meaningful, that is, large (small) distortion should correspond to bad (good) subjective quality. (2) It should be mathematically tractable so as to allow theoretical analyses. (3) It should be computable so that the distortions resulting in an actual system can be determined. Historically the mean-squared error between waveforms has been greatly used because it met attributes (2) and (3), but it has a major drawback of not being sufficiently subjectively meaningful -- especially in systems such as Linear Predictive Coded (LPC) systems. An intuitive explanation for the subjective

inadequacy of mean-squared error is that an ear needs to only recognize the random process producing the waveform to within some accuracy and does not need to accurately reproduce the specific waveform itself, e.g., a "shh" sound is essentially a white noise process and any waveform "typical" or "representative" or "generic" of this process (in the sense of the ergodic theorem) will sound the same, even though such waveforms may differ drastically in individual appearance and hence in mean-squared error. Demanding a small mean-squared error in a speech compression system will therefore often require far more bits and much more accuracy than the human ear requires.

The highly successful LPC systems, however, model speech as a composite or "switched" source formed by outputting segments of stationary and ergodic subsources for intervals of time that are long enough for an observer to estimate the process (or model of a process) being observed and then to transmit a description of the process rather than the actual observed waveform. To measure the distortion of such a system one is naturally led to a distortion measure that measures the closeness of the original and reproduced processes or models rather than the actual waveforms, e.g., between the power spectral densities or related quantities.

To compute the distortion one must view the actual waveforms and estimate such power spectral densities via time-average correlation and Fourier transforms, spectrum analyzers, or appropriately "smoothed" estimates. For a discussion of estimating spectra from waveforms see, e.g., Brillinger [1]. Here it is assumed that the time windows are long enough for the ergodic theorem to ensure that these sample averages nearly equal their expectations, the "true" power spectral densities

involved. Thus, such distortion measures can be viewed as measures of the distortion between power spectral densities or related second order properties of two processes rather than distortion between waveforms. A general discussion and motivation along with relevant references for such distortion measures may be found in Gray and Markel [2]. and related discussions may be found in Viswanathan, et. al. [3] and Makhoul [4].

In addition, some of these distortion measures on spectra or models have proved amenable to analysis, permitting the development of subjectively meaningful mathematical bounds on optimal performance in LPC coded speech with single-symbol quantization of the reflection coefficients [5,6]. These preliminary results suggest that more general techniques from information and communications theory may be applicable to obtain useful performance bounds on more general speech communication systems, for example, LPC systems followed by data compression systems with memory.

Many of these measures appear quite different, possess different properties, and have proved useful for different applications. In [2], Gray and Markel develop some properties and interrelations and discuss the application of some of these measures. In this paper we expand their work by developing more of the properties and interrelations of their distortion measures and some other related distortion measures. Of particular interest is the question of when a class of distortion measures is "equivalent" in the sense that good (bad) performance under one class measure means good (bad) performance under any other. This implies that if one member of the class is subjectively meaningful, then so are all of the others and hence a designer can select a distortion measure from the class on the basis of tractability or

computational efficiency for a particular problem. A second goal of this paper is to provide some interpretations and implication of these distances for coding, prediction, and detection theory that were not given in [2]. These properties help to provide some mathematical intuition as to why these measures are subjectively useful.

2. PRELIMINARIES

L_p Spaces (see, e.g., Ash [7], Ch. 2)

Let $\mathfrak{M}$ denote the space of all measurable complex-valued functions f on $[-\pi,\pi]$. For $p \geq 1$ define L_p as the subspace of $\mathfrak{M}$ containing all f for which

$$\|f\|_p \triangleq \{(2\pi)^{-1}\int_{-\pi}^{\pi} |f(\theta)|^p d\theta\}^{1/p} < \infty \quad ,$$

that is, the integral exists and is finite. If we consider f and g to be equal if

$$\int_{\theta:f(\theta)\neq g(\theta)} d\theta = 0 \quad ,$$

(f=g almost everywhere), then L_p is a normed linear space with norm $\|\cdot\|_p$. The L_p norms are successively stronger in the sense that

$$\|f\|_p \leq \|f\|_q \quad \text{if } q \geq p \quad . \tag{2.1}$$

Important inequalities are the Minkowski or triangle inequality

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p \quad ,$$

the implied inequality

$$\|f-g\|_p \geq \left| \|f\|_p - \|g\|_p \right| \quad , \tag{2.2}$$

Holder's inequality

$$\|fg\|_1 \leq \|f\|_p\|g\|_q \quad , \quad (1/p) + (1/q) = 1,$$

and its special case the Cauchy-Schwartz inequality

$$\|fg\|_1 \leq \|f\|_2\|g\|_2 \quad .$$

Random Processes (Grenader and Szegö [8], Ch. 10; Doob [9], Ch. 10 & 12)

Let $\{X_n\}_{n=-\infty}^{\infty}$ be a real-valued zero mean wide-sense stationary discrete-time random process. We consider discrete time both for simplicity and to be consistent with the speech literature which focuses on sampled waveforms (assumed sampled at a sufficiently high rate so that negligible distortion occurs). Define the autocorrelation function $r(k) = E\, X_n X_{n-k}$, where E denotes expectation, and assume that

$$\sum_{n=-\infty}^{\infty} |r(n)| < \infty \tag{2.3}$$

so that the power spectral density

$$f(\theta) = \sum_{n=-\infty}^{\infty} r(n) e^{-in\theta}, \quad \theta \in [-\pi,\pi]$$

is well-defined, continuous, even, and $f \in L_1$ (in fact, (2.3) implies that f is bounded). Furthermore,

$$r(n) = (2\pi)^{-1} \int_{-\pi}^{\pi} e^{in\theta} f(\theta)\, d\theta \quad . \tag{2.4}$$

When we begin with a spectral density and form r as above, we will often denote it by $r_f(n)$.

By standard arguments from the theory of random processes we can go the other way, that is, given a nonnegative real-valued even $f \in L_1$ there exists a random process $\{X_n\}$ having f as spectral density and $r(n)$ of (2.4) as an autocorrelation function. Thus we can define the space $\mathcal{S}$ of all power spectral densities as the collection of all real-valued, nonnegative, even $f \in L_1$.

A process $\{X_n\}$ is said to be white if $E\, X_n = 0$ and $r(n) = r(0)\delta_n$ ($\delta_n = 1$ for $n = 0$, 0 otherwise) in which case $f(\theta) = r(0)$, $\theta \in [-\pi,\pi]$.

Linear Filters

A (causal and time-invariant) linear filter is described by a δ-response (response to a Kronekker δ) h_k; $k=0,1,\ldots$ If the filter input is $\{X_n\}$, then the output process $\{Y_n\}$ is described by the discrete convolution

$$Y_n = \sum_{k=0}^{\infty} h_k X_{n-k}$$

where the sum exists as a limit in the mean if

$$\sum_{k=0}^{\infty} h_k^2 < \infty \quad .$$

If

$$\sum_{k=0}^{\infty} |h_k| < \infty$$

the filter is said to be stable. A filter is also described by its transfer function $H(e^{i\theta})$ where

$$H(z) = \sum_{k=0}^{\infty} h_k z^{-k} \ .$$

We define h_0 as the gain of the filter. If $h_0 = 1$, the filter is called monic. Any filter can be written as the cascade of a monic filter and a gain. Both H and h will be referred to as filters. Given two filters h and g, the cascade filter d (or D) is defined by the convolution

$$d_n = \sum_{k=0}^{\infty} h_k g_{n-k}$$

or $D(z) = H(z)G(z)$. Note that since the filters are causal, we have that

$$d_0 = h_0 g_0 \quad . \tag{2.5}$$

If g is such that $d_n = \delta_n$ (or, equivalently, $H(z)G(z) = 1$) we say that g is the inverse filter of h (or vice-versa). Note that if h and g are inverse filters, then (2.5) implies that

$$h_0 g_0 = 1 \quad . \tag{2.6}$$

If a random process $\{X_n\}$ with spectral density f is input to a filter H, then the output process has power spectral density $f(\theta)|H(e^{i\theta})|^2$. In particular, if the input process is white with $f(\theta) = r(0) = \sigma^2$, then the output process has power spectral density $\sigma^2|H(e^{i\theta})|^2$. The spectral factorization theorem states that all non-deterministic processes have a second order model of this form with H monic. A process $\{X_n\}$ with spectral density f is nondeterministic if

$$\int_{-\pi}^{\pi} \ln f(\theta)d\theta > -\infty \quad . \tag{2.7}$$

To be precise, the spectral factorization theorem states that $\{X_n\}$ is nondeterministic if and only if the spectral density f has the following form:

$$f(\theta) = |f^+(\theta)|^2 \quad ,$$

where

$$f^+(\theta) = \sigma_f B(e^{i\theta})$$

$$B(z) = \sum_{k=0}^{\infty} b_k z^{-k} \neq 0, \quad |z| > 1$$

$$b_0 = 1 \quad ,$$

$$\sum_{k=0}^{\infty} |b_k|^2 < \infty \quad ,$$

$$\sigma_f^2 = \exp((2\pi)^{-1} \int_{-\pi}^{\pi} \ln f(\theta)d\theta) \quad ,$$

that is, $B(z)$ is analytic in the open unit circle, B is a monic filter, and f^+ is a causal filter with gain σ_f. Intuitively, a process is nondeterministic if and only if it can be represented to second order by a one-sided or causal moving average

$$X_n = \sigma_f \sum_{k=0}^{\infty} b_k Z_{n-k}$$

where $\{Z_n\}$ is white with $E Z_n^2 = 1$. The white process $\{\sigma_f Z_n\}$ is called the innovations process of $\{X_n\}$. In other words, a white process which drives a causal monic filter to produce $\{X_n\}$ is the innovations of $\{X_n\}$.

We make for later convenience the assumption that $1/f \in L_1$. This implies (2.7) from Jensen's inequality.

By assumption $f \in L_1$ and hence from Jensen's inequality we have

$$\sigma_f^2 \leq (2\pi)^{-1} \int_{-\pi}^{\pi} f(\Theta) d\Theta = r(0) < \infty \tag{2.8}$$

with equality if and only if $f(\Theta) = \sigma_f^2 = r(0)$, $\Theta \in [-\pi,\pi]$, i.e., the process is white and equals its own innovations. Since $\sigma_f^2 < \infty$, $f(\theta)$ can also be expressed in factored form with

$$f^+(\Theta) = \sigma_f/C(e^{i\Theta}) \quad ,$$

$$C(z) = \sum_{k=0}^{\infty} c_k z^{-k} \neq 0 \quad , \quad |z| > 1 \quad ,$$

$$c_0 = 1 \quad ,$$

$$\sum_{k=0}^{\infty} |c_k|^2 < \infty \quad ,$$

with σ_f as before. This yields a one-sided autoregressive second order

model of the form

$$X_n = \sum_{k=1}^{\infty} c_k X_{n-k} + \sigma_f Z_n \quad .$$

Comparison with the moving average model shows that B and C are inverse filters.

Denote by $\mathcal{N}$ the class of spectral densities (members of $\mathcal{S}$) for which $1/f \in L_1$. Thus if $f \in \mathcal{N}$ it is nondeterministic (and so is $1/f$) and has both moving average and autoregressive models. Note also that $\sigma_{1/f} = 1/\sigma_f$.

A crucial facet of the preceeding models is their causality. Any power spectral density $f \in \mathcal{S}$ can be modeled to second order as the output of the noncausal filter $f^{1/2}$ (positive square root) driven by $\{Z_n\}$ yielding a two-sided moving average representation $X_n = \sum_{k=-\infty}^{\infty} b'_k Z_{n-k}$. Hence we will refer to f^+ as a causal model and $f^{1/2}$ as a noncausal model for f. Note that $f^{1/2}$ exists more generally and that if both exist, $f^{1/2} = |f^+|$.

Linear Prediction (Grenander and Szegö [8], Ch. 10, Doob [9], Ch. 12, Gray and Markel [23])

For a finite integer m and a process $\{X_n\}$, form a one-step linear predictor $\hat{X}_n$ of X_n of the form

$$\hat{X}_n = \sum_{k=1}^{m} h_k X_{n-k}$$

with average squared error

$$E(e_n^2) = E((X_n - \hat{X}_n)^2) = E((X_n - \sum_{k=1}^{m} h_k X_{n-k})^2) .$$

Define the monic filter a_k by $a_0 = 1$, $a_k = -h_k$, $k \neq 0$ ($h_k \stackrel{\Delta}{=} 0$ if $k < 0$ or $k > m$). We have then

$$E(e_n^2) = E(|\sum_{k=0}^{m} a_k X_{n-k}|^2) = (2\pi)^{-1}\int_{-\pi}^{\pi} f(\Theta)|A(e^{i\Theta})|^2 d\Theta , \qquad (2.9)$$

where $A(z) = 1-H(z) = 1-\sum_{k=1}^{m} h_k z^{-k}$. We wish to find the monic A (and hence H) that minimizes $E(e_n^2)$ with the resulting squared error σ_m^2. The solution is well-known to be given by the A, say $\hat{A}$, that solves the linear system of equations

$$\sigma_m^2 = (2\pi)^{-1}\int_{-\pi}^{\pi} f(\Theta)\hat{A}(e^{-i\Theta})d\Theta = \sum_{k=0}^{m} \hat{a}_k r_f(k)$$

$$0 = (2\pi)^{-1}\int_{-\pi}^{\pi} f(\Theta)\hat{A}(e^{-i\Theta})e^{-ij\Theta}d\Theta = \sum_{k=0}^{m} \hat{a}_k r_f(k-j), \qquad (2.10)$$

$$j = 1,2,\ldots,m$$

Note this is easily proved by observing that for any monic m^{th} order filter G (2.10) implies

$$(2\pi)^{-1}\int_{-\pi}^{\pi} f(\Theta)\hat{A}(e^{-i\Theta})G(e^{+i\Theta})d\Theta = \sigma_m^2 + \sum_{k=1}^{m} g_k (2\pi)^{-1}\int_{-\pi}^{\pi} f(\Theta)\hat{A}(e^{-i\Theta})e^{+ik\Theta}d\Theta = \sigma_m^2 \qquad (2.11)$$

and hence for any m^{th} order monic filter A we have from (2.9) that

$$E(e_n^2) = (2\pi)^{-1}\int_{-\pi}^{\pi} f(\Theta)|A(e^{i\Theta})|^2 d\Theta$$

$$= (2\pi)^{-1}\int_{-\pi}^{\pi} f(\Theta)|\hat{A}(e^{i\Theta}) + A(e^{i\Theta}) - \hat{A}(e^{i\Theta})|^2 d\Theta$$

$$= \sigma_m^2 + (2\pi)^{-1}\int_{-\pi}^{\pi} f(\Theta)|A(e^{i\Theta}) - \hat{A}(e^{i\Theta})|^2 d\Theta$$

$$+ 2\mathcal{R}_e\{(2\pi)^{-1}\int_{-\pi}^{\pi} f(\Theta)\hat{A}(e^{-i\Theta})(A(e^{i\Theta})-\hat{A}(e^{i\Theta}))d\Theta\}$$

$$= \sigma_m^2 + (2\pi)^{-1}\int_{-\pi}^{\pi} f(\Theta)|A(e^{i\Theta}) - \hat{A}(e^{i\Theta})|^2 d\Theta \geq \sigma_m^2 \qquad (2.12)$$

with equality if and only if $A = \hat{A}$ (almost everywhere). This is simply the orthogonality principal.

It is well-known that the system of equations (2.10) is equivalent to the system of "correlation matching" equations [10,4]

$$r_f(k) = (2\pi)^{-1}\int_{-\pi}^{\pi} f(\Theta)e^{ik\Theta}d\Theta = (2\pi)^{-1}\int_{-\pi}^{\pi}\{\sigma_m^2/|\hat{A}(e^{i\Theta})|^2\}e^{ik\Theta}d\Theta\ , \qquad (2.13)$$

$$k = 0,1,\ldots,m$$

That (2.13) implies (2.10) is easy, the converse implication seems more difficult to prove. Note that (2.13) implies that for any m^{th} order filter $G(\Theta) = \sum_{k=0}^{m} g_k e^{-ik\Theta}$ we have

$$(2\pi)^{-1}\int_{-\pi}^{\pi} f(\Theta)|G(\Theta)|^2 d\Theta = (2\pi)^{-1}\int_{-\pi}^{\pi} d\Theta|G(\Theta)|^2\sigma_m^2/|\hat{A}(e^{i\Theta})|^2 \qquad (2.14)$$

The minimum value σ_m^2 of $E(e_n^2)$ is given by

$$\sigma_m^2 = \frac{\det(T_{m+1}(f))}{\det(T_m(f))}$$

where $T_m(f) = \{(2\pi)^{-1} \int_{-\pi}^{\pi} f(\theta)e^{i(k-j)\theta}d\theta;\ k,j=0,1,\ldots,m-1\} =$ $\{r_f(k-j);\ k,j=0,1,\ldots,m-j\}$ is the m^{th} order Toeplitz matrix of the spectral density f. It is well-known from the theory of Toeplitz forms (Grenander and Szegö [8]) that $\sigma_m^2 \to \sigma_f^2$ as $m \to \infty$ and that $\sigma_m^2 \geq \sigma_f^2$. A similar argument to the preceeding shows that if $m = \infty$, then $\sigma_\infty^2 = \sigma_f^2$ whence $\sigma_m^2 \geq \sigma_f^2$ for all m. In this case, if f has autoregressive model $\sigma_f^2/|A|^2$, then $\hat{A} = A$ is the best predictor and the resulting e_n is white with $E\,e_n^2 = \sigma_f^2$, that is, passing $\{X_n\}$ through the filter A yields its innovations process and hence A is called a whitening filter for f, ($1/f^+$ and $1/f^{1/2}$ also whiten f).

3. SPECTRAL RATIO DISTORTION MEASURES

In this section we consider basic properties of distortion measures on $\mathcal{N}$. In the next section several specific examples are introduced and compared.

A distortion measure is a generalization of the notion of a distance or metric. A distortion measure $d(\cdot,\cdot)$ on $\mathcal{N}$ is simply an assignment of a nonnegative extended real number to each pair f, g in $\mathcal{N}$. Intuitively, $d(f,g)$ represents the distortion or cost or "badness of approximation" of reproducing f as g. Without loss of generality we can assume that $d(f,f) = 0$ (see, e.g., Berger [11]).

Distortion measures may or may not have the following properties: A distortion measure is (a) symmetric if $d(f,g) = d(g,f)$, all $f,g \in \mathcal{N}$; (b) finite-valued if $d(f,g) < \infty$, all $f,g \in \mathcal{N}$; (c) positive definite if $d(f,g) = 0$ means $f = g$ (almost everywhere); (d) metric (actually, pseudo-metric) if $d(f,g) \le d(f,h) + d(h,g)$, all $f,g,h \in \mathcal{N}$. A distortion measure is called a distance or a metric if it has all of these properties. Metrics have additional structure over general distortion measures, but most basic theoretical results for distortion measures such as information theoretic optimal performance bounds do not require (a), (c) or (d). They do require (b) (at least with probability one since communication with finite average distortion is otherwise impossible). In particular, nonsymmetric distortion measures may not be as easy to work with, but they have no inherent mathematical drawbacks to communications theory and may in fact be more appropriate for certain situations. The metric property, however, is quite useful since it allows us to conclude that if in a given communications system

the reproduction is produced in two steps and each step results in small distortion, then the overall distortion will also be small. Some of the distortion measures here will have similar properties since they are defined in terms of L_p norms.

Given two distortion measures d_1 and d_2 on a common space $\mathcal{T}$, we shall be concerned with which is "stronger" or "weaker." We say that d_1 is stronger than d_2 or implies d_2 and write $d_1 \Rightarrow d_2$ if small enough distortion under d_1 implies that d_2 is also small, that is, given $f \in \mathcal{T}$ and $\epsilon > 0$ there is a $\delta > 0$ such that if $d_1(f,g) \le \delta$, then $d_2(f,g) \le \epsilon$. If $d_1 \Rightarrow d_2$ and $d_2 \Rightarrow d_1$, we say that d_1 and d_2 are equivalent and write $d_1 \Leftrightarrow d_2$. Intuitively, equivalent distortion measures yield the same notions of "good" and "bad" performance even though their numerical values may differ. For example, $\|f-g\|_2$ (which is a metric) and $\|f-g\|_2^2$ (which is not) are obviously equivalent distortion measures. Clearly this is actually an equivalence relation in the sense that $d_1 \Leftrightarrow d_2$ and $d_2 \Leftrightarrow d_3$ implies $d_1 \Leftrightarrow d_3$. The intuitive importance of equivalence lies in the fact that if a distortion measure d is subjectively meaningful, then so are all other distortion measures equivalent to d since small and large values yield the same notion of "good" and "bad," only the numerical requirements of small and large change.

In some cases it is useful to define distortion measure in terms of other distortion measures. For example, given distortion measures $d_1(f,g)$ and $d_2(f,g)$ on $\mathcal{T}$ one can define

$$d^{(q)}(f,g) = (d_1(f,g)^q + d_2(f,g)^q)^{1/q} \tag{3.1}$$

where $q \geq 1$ is a parameter. Sometimes a scaling is also included. The special cases $q = 1$ and 2 are the most common, but general q are sometimes considered as most of the properties remain true and a greater variety of distortion measures is thereby included. The parameter q can be chosen for convenience since, as we now show, the distortion measures $d^{(q)}(f,g)$ are all equivalent for fixed d_1 and d_2. From Ash [7], pp. 83-88, we have that

$$a^q + b^q \leq (a+b)^q \leq 2^{q-1}(a^q + b^q) \quad , \quad a,b \geq 0, \; q \geq 1 \quad , \tag{3.2}$$

and hence

$$d^{(q)}(f,g) \leq d_1(f,g) + d_2(f,g) \leq 2^{1-1/q} d^{(q)}(f,g) \tag{3.3}$$

and therefore $d^{(q)}(f,g) \Leftrightarrow d^{(1)}(f,g)$ for all q. Note that $d^{(q)}(f,g) \Rightarrow d_i(f,g)$, $i = 1,2$, but that it may not be true, for example, that $d_1(f,g) \Rightarrow d^{(q)}(f,g)$.

An example of the previous construction is to symmetrize a nonsymmetric distortion measure. A nonsymmetric distortion measure can be symmetrized in a number of ways. The most common is to define $d_1(f,g) = d(f,g)$ and $d_2(f,g) = d(g,f)$ and use (3.1), that is, to define

$$d^{(q)}(f,g) \stackrel{\Delta}{=} (d(f,g)^q + d(g,f)^q)^{1/q} \tag{3.4}$$

Equation (3.3) implies that the $d^{(q)}$ are equivalent for all q.

A common useful class of distortion measures are difference distortions having the following form: $d(f,g)$ is a difference distortion measure on $\mathcal{M} \subset L_p$ if there is a function $\varphi:(-\infty,\infty) \to [0,\infty)$ such that

$\varphi(0) = 0$, $\varphi(x) \geq 0$, and

$$d(f,g) = \|\varphi(f-g)\|_p \quad , \tag{3.5}$$

where $f-g$ is well-defined since f and g are members of a linear space. It is usually assumed that $\varphi(|x|)$ is nondecreasing with $|x|$ or, more strongly, that $\varphi(x)$ is convex $\cup$ (for example, $\varphi(|x|) = |x|^q$, $q \geq 1$). An alternate class of distortion measures sometimes referred to as difference distortion measures reverses the roles of norm and φ and sets

$$d'(f,g) = \varphi(\|f-g\|_p) \quad , \tag{3.6}$$

with φ having the above properties. We shall call this a norm-difference distortion measure.

Most distortion measures arising in speech applications, however, are not difference measures. Instead they are ratio distortion measures having the following form: Let $\varphi:[0,\infty) \to [0,\infty)$ satisfy $\varphi(1) = 0$, $\varphi(x) \geq 0$, and $\varphi(x)$ is a convex $\cup$ function of x (with a minimum now at 1 instead of 0). A distortion measure of the form

$$d_\varphi(f,g) = \|\varphi(f/g)\|_p$$

is called a ratio distortion measure on $\mathcal{H}$. The subscript φ will often be replaced by a mnemonic. We also consider gain-normalized distortion measures of the form

$$d_{n\varphi}(f,g) = d_\varphi(f/\sigma_f^2, g/\sigma_g^2) = \left\|\varphi\left(\frac{f/\sigma_f^2}{g/\sigma_g^2}\right)\right\|_p ,$$

where the subscript "n" is an abbrevation for "normalized." Note that a ratio distortion measure can also be considered as a difference

distortion measure on the space $\tilde{\mathcal{M}}$ of all functions of the form $\ell n\ f$, $f \in \mathcal{M}$, but we work with $\mathcal{M}$ as our basic space as it contains the basic structural properties of spectra. Analagous to (3.6) we can also have a norm-ratio distortion measure of the form

$$d'_\varphi(f,g) = \varphi(\|f/g\|_p)$$

Note that for $p = 1$ we have from Jensen's inequality that

$$d'_\varphi(f,g) = \varphi(\|f/g\|_1) \le \|\varphi(f/g)\|_1 = d_\varphi(f,g) \tag{3.7}$$

and therefore for L_1 norms

$$d_\varphi \Rightarrow d'_\varphi \tag{3.8}$$

A variation on the ratio distortion measure that occurs in speech processing is the gain-optimized distortion measure. Here we begin with a ratio distortion measure d_φ, but the dependence on the reproduction gain σ_g^2 is removed by replacing it with a gain σ^2 that minimizes d_φ. This is usually done for one of two reasons. First, we may ignore the original gain of a reproduction symbol and replace it by a gain chosen to minimize the given distortion measure. Second, by removing dependence of the distortion measure d_φ on a reproduction parameter such as the gain it allows us the freedom of using a different distortion measure on the gains. We thus define a gain-optimized distortion measure

$$d_\varphi^o(f,g) = \inf_{\sigma^2 \ge 0} d_\varphi(f, \sigma^2 g/\sigma_g^2) \quad .$$

If the infimum is a minimum, the optimum σ^2 is denoted σ_o^2 and called the optimum reproduction gain. Note that we remove the original reproduction gain by normalizing g and replace it by the new gain σ^2,

but since the distortion measure is a ratio measure we can also write

$$d_{\varphi}^{o}(f,g) = \inf_{\sigma^2 \geq 0} \left\| \varphi\left(\frac{f/\sigma^2}{g/\sigma_g^2}\right) \right\|_p$$

$$= \inf_{\sigma^2 \geq 0} d_{\varphi}(f/\sigma^2, g/\sigma_g^2)$$

Note that obviously

$$d_{\varphi}(f,g) \geq d_{\varphi}^{o}(f,g) \tag{3.9}$$

and therefore

$$d_{\varphi}(f,g) \Rightarrow d_{\varphi}^{o}(f,g) \quad . \tag{3.10}$$

Given two spectral ratio distortion measures $d_1(f,g) = \|\varphi_1(f/g)\|_p$ $d_2(f,g) = \|\varphi_2(f/g)\|_p$, we can form a new distortion measure $d^{(q)}(f,g)$ as in (3.1) by

$$d^{(q)}(f,g) = (\|\varphi_1(f/g)\|_p^q + \|\varphi_2(f/g)\|_p^q)^{1/q} \quad . \tag{3.11}$$

Alternatively, we might combine the φ functions before taking the norm, for example forming

$$\varphi^*(f/g) = \frac{1}{2}(\varphi_1(f/g) + \varphi_2(f/g))$$

(the 1/2 is for convenience) and then define

$$d^*(f,g) = \|\varphi^*(f/g)\|_p$$

$$= \left\|\frac{1}{2}(\varphi_1(f/g) + \varphi_2(f/g))\right\|_p$$

$$= \frac{1}{2}\|\varphi_1(f/g) + \varphi_2(f/g)\|_p \tag{3.12}$$

The measures of (3.11) and (3.12) are related for $q = 1$ by

$$\frac{1}{2}\, d^{(1)}(f,g) = \frac{1}{2}(\|\varphi_1(f/g)\|_p + \|\varphi_2(f/g)\|_p) \le \max(\|\varphi_1(f/g)\|_p,\ \|\varphi_2(f/g)\|_p) \le$$

$$\|\varphi_1(f/g)+\varphi_2(f/g)\|_p = d^*(f,g) \le \|\varphi_1(f/g)\|_p + \|\varphi_2(f/g)\|_p = d^{(1)}(f,g),$$ and

hence since $d^{(q)} \Leftrightarrow d^{(1)}$ we have that

$$d^* \Leftrightarrow d^{(q)} \quad , \quad q = 1,2,\ldots \tag{3.13}$$

that is, both forms of symmetrized measures are equivalent. In particular, if one wishes to symmetrize a nonsymmetric distortion measure $\|\varphi(f/g)\|_p$, then

$$(\|\varphi(f/g)\|_p^q + \|\varphi(g/f)\|_p^q)^{1/q} \Leftrightarrow \frac{1}{2}\, \|\varphi(f/g) + \varphi(g/f)\|_p \ . \tag{3.14}$$

Obviously other alternatives to (3.12) exist for forming φ^* from φ_1 and φ_2, e.g., $\varphi^{**}(f/g) = \{\varphi_1(f/g)\varphi_2(f/g)\}^{1/2}$ yields another ratio distortion measure. The arithmetic mean of (3.12), however, seems the most useful.

Another form of implication and equivalence of distortion measures is the following: We say that d_1 is stronger in a coding sense than d_2 and write $d_1 \supset d_2$ if for each $f,g,g' \in \eta$ we have that $d_1(f,g) \le d_1(f,g')$ implies that $d_2(f,g) \le d_2(f,g')$, i.e., if g' is a worse reproduction of f than g is under d_1, then it is also worse under d_2. If $d_1 \supset d_2$ and $d_1 \subset d_2$, we write $d_1 \subset\supset d_2$ and say that d_1 and d_2 are coding equivalent. The name and application of this concept arises in the following coding or quantization problem. Consider $f \in \eta$ as a symbol in an alphabet η and let $\hat{\eta}$ be a subset of η called a reproduction space or codebook (usually $\hat{\eta}$ has a finite number of members). Given a distortion measure d_i, define the minimum distortion quantizer (or coder) $\hat{f}_i : \eta \to \hat{\eta}$ by

$$\hat{f}_i(f) = g \in \hat{\eta} \text{ if } d_i(f,g) \le d_i(f,g') , \text{ all } g' \in \hat{\eta} ,$$

with some tie-breaking rule. Thus $\hat{f}_i$ picks the closest or minimum distortion (under d_i) reproduction symbol to f. If $d_1 \supset d_2$, then

$$d_2(f,\hat{f}_2(f)) = d_2(f,\hat{f}_1(f)) ,$$

that is, a closest reproduction symbol $\hat{f}_1(f)$ under d_1 is also a closest reproduction under d_2. (It may or may not be true that $\hat{f}_1(f) = \hat{f}_2(f)$ depending on the tie breaking rule). If $d_1 \subset\supset d_2$, then the tie-breaking rules can be chosen so that $\hat{f}_1(f) = \hat{f}_2(f)$, that is, coding equivalent distortion measures result in the same code, Note, however, that the code may be "good" under one distortion measure yet "bad" under another in the sense of average performance.

4. EXAMPLES

In this section several examples of spectral ratio distortion measures introduced in the speech literature along with some other related measures are defined and motivated. In the next section their properties and interrelations are developed. We begin by listing the various measures along with comments on each. In each case we set $d(f,g) = \infty$ if the given integral does not exist.

1) The Itakura-Saito Distortion Measure

$$d_{IS}(f,g) = \|f/g - 1 - \ell n(f/g)\|_1 \tag{4.1}$$

This is a ratio distortion measure with $\varphi(x) = x-1-\ell n\ x$. This distance was introduced by Itakura and Saito [12] and has the property that for fixed f and a class $\mathcal{N}_m \overset{\Delta}{=} \{$all $g \in \mathcal{N}$ such that $g(\theta) = \sigma^2/|\sum_{k=0}^{\infty} a_k e^{-ik\theta}|^2$, $a_0 = 1\}$, then the $g \in \mathcal{N}_m$ minimizing $d_{IS}(f,g)$ is $g(\theta) = \sigma_m^2/|\hat{A}(e^{i\theta})|^2$, where $\hat{A}$ is defined by (2.10) and yields the minimum prediction error over all m^{th} order prediction filters $H = 1-A$. Itakura and Saito also showed that if the underlying process was assumed Gaussian, then minimizing $d_{IS}(f,\sigma^2/|A|^2)$ is approximately equivalent to finding a maximum likelihood guess of A given a sample power spectral density f. A related less known property of this measure under a Gaussian assumption is the following: If f and g are power spectral densities of two zero mean Gaussian processes, then let $p_f^n(x^n)$ and $p_g^n(x^n)$, $x^n \in (-\infty,\infty)^n$, denote the resulting probability density functions and define the n^{th} order relative entropy (or Kullback-Leibler number or directed divergence) [13,14,15]

$$I_n(f,g) = \int dx^n p_f^n(x^n) \ell n(p_f^n(x^n)/p_g^n(x^n)) \quad .$$

This quantity is used both in detection and information theory.

It is well-known [13,14,15] that for Gaussian processes

$$I_n(f,g) = \frac{1}{2} \ln \det \frac{T_n(g)}{T_n(f)} + \frac{1}{2} \operatorname{tr} T_n(f) T_n(g)^{-1} - \frac{n}{2} , \qquad (4.2)$$

where "tr" denotes the trace of a matrix and, as previously discussed, $T_n(f)$ is the n^{th} order autocorrelation matrix of f. From the asymptotic eigenvalue theorem for Toeplitz matrices [8,16,17], the normalized directed divergence has limit

$$\begin{aligned} I(f,g) &= \lim_{n \to \infty} n^{-1} I_n(f,g) \\ &= \frac{1}{2} \ln \frac{\sigma_g^2}{\sigma_f^2} + \frac{1}{2} (2\pi)^{-1} \int_{-\pi}^{\pi} \frac{f(\theta)}{g(\theta)} d\theta - \frac{1}{2} \\ &= \frac{1}{2} d_{IS}(f,g) \qquad , \qquad (4.3) \end{aligned}$$

that is, the Itakura-Saito distance between f and g is exactly half the asymptotic per symbol Kullback-Leibler number under a Gaussian assumption.

We note that

$$\begin{aligned} d_{IS}(f,g) &= (2\pi)^{-1} \int_{-\pi}^{\pi} \left(\frac{f(\theta)}{g(\theta)} - 1 - \ln \frac{f(\theta)}{g(\theta)} \right) d\theta \\ &= (2\pi)^{-1} \int_{-\pi}^{\pi} \frac{f(\theta)}{g(\theta)} - 1 - \ln \frac{\sigma_f^2}{\sigma_g^2} \\ &= r_{f/g}(0) - 1 - \ln \frac{\sigma_f^2}{\sigma_g^2} \qquad (4.4) \end{aligned}$$

where the integrand in the leftmost integral is nonnegative from the inequality $\ln x \leq x-1$.

2) <u>The Itakura Distortion Measure</u>

$$d_I(f,g) = \ell n \left\{ (2\pi)^{-1} \int_{-\pi}^{\pi} \frac{f(\theta)/\sigma_f^2}{g(\theta)/\sigma_g^2} d\theta \right\}$$

$$= \ell n(\sigma_g^2 r_{f/g}(0)/\sigma_f^2) \quad . \tag{4.5}$$

This distortion was introduced by Itakura [18] as the gain-optimized Itakura-Saito distortion, that is,

$$d_I(f,g) = d_{IS}^o(f,g) \tag{4.6}$$

The Itakura distortion is a gain-normalized norm-ratio distortion measure with $\varphi(x) = \ell n\ x$ for $x \geq 1$ since

$$d_I(f,g) = \ell n \left(\left\| \frac{f/\sigma_f^2}{g/\sigma_g^2} \right\|_1 \right) \tag{4.7}$$

where the argument is greater than 1 from (2.8) applied to f/g coupled with the fact that $\sigma_{f/g}^2 = \sigma_f^2/\sigma_g^2$.

3) <u>Model Distortion Measures</u>

Define the causal model (or filter) distortion measure

$$d_{cm}(f,g) = \|1-f^{+}/g^{+}\|_2 \qquad , \qquad (4.8)$$

and the gain-normalized causal model distortion

$$d_{ncm}(f,g) = d_{cm}(f/\sigma_f^2,\ g/\sigma_g^2)$$

$$= \left\| 1 - \frac{f^{+}/\sigma_f}{g^{+}/\sigma_g} \right\|_2 \qquad (4.9)$$

The gain-normalized causal model distortion measure is a gain-normalized spectral ratio distortion measure (also a norm-spectral measure since

$$d_{ncm}(f,g)^2 = 1-2\Re_e\left\{(2\pi)^{-1}\int_{-\pi}^{\pi}\frac{f^{+}(\theta)/\sigma_f^2}{g^{+}(\theta)/\sigma_g^2}\,d\theta\right\}$$

$$+ (2\pi)^{-1}\int_{-\pi}^{\pi}\frac{f(\Theta)/\sigma_f^2}{g(\Theta)/\sigma_g^2}\,d\Theta = (2\pi)^{-1}\int_{-\pi}^{\pi}\frac{f(\Theta)/\sigma_f^2}{g(\Theta)/\sigma_g^2}\,d\Theta-1$$

$$= \sigma_g^2 r_{f/g}(0)/\sigma_f^2 - 1$$

$$= e^{d_I(f,g)} - 1 \qquad , \qquad (4.10)$$

where we have used the fact that the f^{+}/σ_f and σ_g/g^{+} are monic and causal (from (2.5)) and hence from (2.6) the bracketed term above is 1. Note that

$$d_I(f/\sigma_f^2,\ g/\sigma_g^2) \quad \Leftrightarrow \quad d_{ncm}(f,g) \qquad .$$

The gain-normalized causal model distortion measure was introduced by Itakura [18] as an approximation to the Itakura distortion measure for small values since from (4.10) we have that for small $d_I(f,g)$

$$d^2_{ncm}(f,g) \cong d_I(f,g) \quad .$$

Regardless of approximation, however, (4.10) implies (directly or from the $\ln x \le x-1$ inequality) that

$$d_{ncm}(f,g)^2 \ge d_I(f,g) \tag{4.11}$$

The distortion measure d_{ncm} has the property of the Itakura-Saito distance that for fixed f and the class $\hat{\mathcal{M}}_m$, a minimum distortion $g \in \hat{\mathcal{M}}_m$ will have the form $g(\theta) = \sigma^2/|\hat{A}(e^{i\theta})|^2$, where $\hat{A}$ is defined by (2.10), but σ^2 is arbitrary.

Chaffee [19] also used the gain-normalized causal model distortion measure in his coding (or rate-distortion) approach to speech compression where he used the coding or quantization approach previously described to select a monic filter reproduction and an alternate criterion to select the gain.

The causal model distortion measure is a slight generalization of d_{ncm} and is introduced for comparison and interpretation purposes. Note that analagous to (4.10)

$$\begin{aligned} d_{cm}(f,g)^2 &= 1 + (2\pi)^{-1}\int_{-\pi}^{\pi}\frac{f(\theta)}{g(\theta)}\,d\theta - 2\mathcal{R}_e\left\{(2\pi)^{-1}\int_{-\pi}^{\pi}\frac{f^+(\theta)}{g^+(\theta)}\,d\theta\right\} \\ &= 1 + r_{f/g}(0) - 2\sigma_f/\sigma_g \end{aligned} \tag{4.12}$$

and hence d_{cm} can be thought of as a gain-biased spectral ratio distortion measure. We can also consider a gain-optimized causal model ratio measure which is easily shown to be

$$\begin{aligned} d^o_{cm}(f,g)^2 &= 1 - \frac{\sigma_f^2/\sigma_g^2}{r_{f/g}(0)} \\ \sigma_0 &= \sigma_g^2 r_{f/g}(0)/\sigma_f \quad . \end{aligned} \tag{4.13}$$

Another related measure is the noncausal model distortion measure

(again introduced for comparison)

$$d_{nm}(f,g)^2 = \|1-(f/g)^{1/2}\|_2^2 = \|1-|f^+/g^+|\|_2^2$$

$$= 1 + r_{f/g}(0) - 2(2\pi)^{-1} \int_{-\pi}^{\pi} \frac{f(\theta)^{1/2}}{g(\theta)^{1/2}}\, d\theta$$

This is a spectral ratio measure with $\varphi(x) = 1-x^{1/2}$.

All the model distortion measures have the following interpretation: Say $\{X_n\}$ has spectral density $f(\Theta) = |F(\Theta)|^2$, where $F(\Theta)$ is the transfer function of a causal model f^+ or noncausal model $|f^+|$. Similarly define $g(\Theta) = |G(\Theta)|^2$ and consider the system of Fig. 1a, where $\{Z_n\}$ is a unit variance zero mean white process. We have that

$$\|1-F/G\|_2^2 = E(Z_n - Y_n)^2 \quad , \tag{4.14}$$

that is, $\|1-F/G\|_2^2$ measures how nearly inverse filters F and $1/G$ are by measuring the average squared error between a white input process and the cascade of F and $1/G$. The closer F and G are, the more "white" the output of the cascade F/G looks since it is close in a squared-error sense to the white input.

Alternatively, consider the system of Fig. 1b. Here $1/F$ is a true whitening filter for $\{X_n\}$ and $1/G$ is a "mismatched" whitening filter. Here again

$$E(Z_n-Y_n)^2 = \|1-F/G\|_2^2$$

and hence $\|1-F/G\|_2^2$ is a measure of the "mismatch" of $1/G$ to F in that it measures the error power between the true whitened process and the mismatch whitened process. This interpretation of d_{ncm} is used by Gray and Markel [2] (wherein $d_{ncm}^2 = \delta/\alpha - 1$).

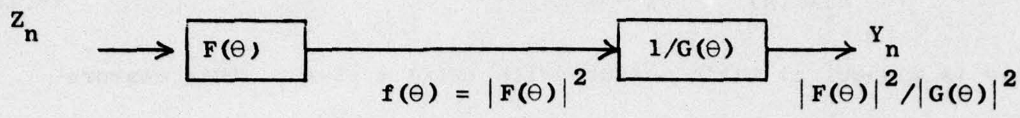

Figure 1a

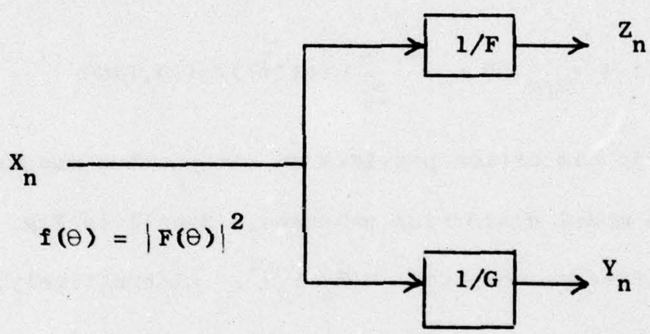

Figure 1b

4) <u>The L_1 Spectral Ratio Distortion Measure</u>

$$d_1(f,g) = \|1 - f/g\|_1 \tag{4.15}$$

This is a spectral ratio measure with $\varphi(x) = |1-x|$. This measure will provide some interesting comparisons with the model distortion measures. We have that

$$\begin{aligned} d_1(f,g) &= (2\pi)^{-1} \int_{-\pi}^{\pi} |1-f(\Theta)/g(\Theta)| d\Theta \\ &= (2\pi)^{-1} \int_{-\pi}^{\pi} (1+f(\Theta)/g(\Theta) - 2\min(f(\Theta)/g(\Theta),1))d\Theta \\ &= 1 + r_{f/g}(0) - \pi^{-1} \int_{-\pi}^{\pi} \min(f(\Theta)/g(\Theta),1)d\Theta \end{aligned} \tag{4.16}$$

The L_1 ratio distortion provides an alternative measure of "mismatch" to the model distortion measures. Recall in Fig. 1 that the model distortion measure computed $E(Z_n - Y_n)^2$. Alternatively, we can measure mismatch by

$$\begin{aligned} |E\, Z_n^2 - E\, Y_n^2| &= |r_{f/g}(0) - 1| \\ &\le \|1 - f/g\|_1 \quad , \end{aligned} \tag{4.17}$$

so that d_1 is an upperbound on the difference of the output powers as opposed to the power of the difference.

All of the preceeding measures are nonsymmetric. We next consider several symmetric distortion measures.

5) Log Spectral Deviation

$$d_{\log}(f,g) = \|\ln f/g\|_p$$

$$= \|\ln f - \ln g\|_p \tag{4.18}$$

The most common choices for p are 2,1, or ∞. This is a spectral ratio measure with $\varphi(x) = |\ln x|$ and is one of the most commonly proposed distortion measures for speech [5, 2, 6]. For $p = 2$ there exist fast techniques for computation of $d_{\log}$ using cepstral approximations, but there do not seem to be fast algorithms for finding the best fit (say in $\mathcal{T}_m$) to a given f. Note that $d_{\log}$ is metric since

$$d_{\log}(f,g) \le d_{\log}(f,g') + d_{\log}(g',g) \quad .$$

The remaining measures are all symmetrized versions of measures 1-4. As was shown in Section 3, there are several equivalent means of symmetrizing measures and hence we can choose the simplest or most useful.

6) The Cosh distortion measure: Symmetrizing the Itakura-Saito distortion as in (3.8) we obtain

$$d^*_{IS}(f,g) = \frac{1}{2}(2\pi)^{-1}\int_{-\pi}^{\pi}\left\{\frac{f(\Theta)}{g(\Theta)} + \frac{g(\Theta)}{f(\Theta)} - 2\right\} d\Theta$$

$$= \frac{1}{2}(2\pi)^{-1}\int_{-\pi}^{\pi}\left\{\frac{f(\Theta)^{1/2}}{g(\Theta)^{1/2}} - \frac{g(\Theta)^{1/2}}{f(\Theta)^{1/2}}\right\} d\Theta$$

$$= \frac{1}{2}\,\|(f/g)^{1/2} - (g/f)^{1/2}\|_2^2 \tag{4.19}$$

We also can write

$$d^*_{IS}(f,g) = \frac{1}{2}\,\|r_{f/g}(0) + r_{g/f}(0) - 2\|_1$$

$$= \frac{1}{2}\{d_{IS}(f,g) + d_{IS}(g,f)\}$$

$$= \frac{1}{2}\,d^{(1)}_{IS}(f,g) \quad ,$$

where $d^{(1)}_{IS}$ is defined by (3.11), that is, both types of symmetrization yield effectively the same measure. This distortion measure was introduced by Gray and Markel [2] and is called the Cosh distortion measure since

$$d_{cosh}(f,g) = d^*_{IS}(f,g) = \|\cosh(\ln f/g) - 1\|_1$$

This measure has some interesting interpretations. First recall the directed divergence discussion of Gaussian processes and the Itakura-Saito distortion. In statistics, detection theory, and information theory [13,14,15] one often uses the symmetrized directed divergence $J = I(1,2)+I(2,1)$, where J is called simply the divergence. Defining $J_n = I_n(1,2)+I_n(2,1)$ we have from (4.2)-(4.3) that

$$\lim_{n\to\infty} n^{-1}J_n = \frac{1}{2}\,d_{IS}(f,g) + \frac{1}{2}\,d_{IS}(g,f)$$

$$= d_{cosh}(f,g) \quad ,$$

that is, the cosh measure is exactly the asymptotic normalized divergence between processes having spectral density f and g under a Gausian assumption.

A second interpretation comes from the theory of random processes. Say we have two stationary Gaussian processes $\{U_n\}$ and $\{V_n\}$ with spectral densities f_1 and f_2 respectively. The squared-error $\bar{\rho}$-distance [20] (or Ornstein distance) between these processes is defined by

$$\bar{\rho}(f_1,f_2) = \inf E((U_0-V_0)^2)$$

where the infimum is over all consistent joint probabilistic descriptions of $\{U_n\}$ and $\{V_n\}$, that is, over all pair processes $\{U_n, V_n\}$ having the original processes as coordinates. Intuitively, given $\{U_n\}$ and $\{V_n\}$, $\bar{\rho}$ measures how well we can "fit" the two processes together in a squared-error sense. For Gaussian process [20]

$$\bar{\rho}(f_1, f_2) = \|f_1^{1/2} - f_2^{1/2}\|_2^2$$

and hence

$$d_{\cosh}(f,g) = \frac{1}{2}\bar{\rho}(f/g, g/f) \quad ,$$

the cosh measure is one-half the $\bar{\rho}$-distance between the "mismatch" whitened process f/g and g/f. Intuitively, instead of comparing f/g to one to see how nearly inverses f and $1/g$ are, we compare f/g to its own inverse g/f. We note that even if the processes are not Gaussian, then $\bar{\rho}(f_1, f_2) \geq \|f_1^{1/2} - f_2^{1/2}\|_2^2$ [20].

7) Gain-Optimized Cosh Measure

As the Itakura distortion was obtained from the Itakura-Saito distance by choosing a reproduction gain, the cosh measure can also be modified in a similar fashion (as suggested by Gray and Markel [2]). Form

$$d_{\cosh}(f, \sigma^2 g/\sigma_g^2) = \frac{1}{2}\{(\sigma_g^2/\sigma^2) r_{f/g}(0) + (\sigma^2/\sigma_g^2) r_{g/f}(0) - 2\}$$

and use calculus to minimize this over σ^2 resulting in

$$\sigma_0^2 = \{r_{f/g}(0)/r_{g/f}(0)\}^{1/2} \tag{4.20}$$

$$\begin{aligned} d^o_{\cosh}(f,g) &= d_{\cosh}(f, \sigma_0^2 g/\sigma_g^2) \\ &= 2(r_{f/g}(0)^{1/2} r_{g/f}(0)^{1/2} - 1) \end{aligned} \tag{4.21}$$

8) Symmetrized Itakura Distortion

$$d_I^*(f,g) = \frac{1}{2}\left(d_I(f,g) + d_I(g,f)\right)$$
$$= \frac{1}{2} \ln r_{f/g}(0) r_{g/f}(0) \quad . \tag{4.22}$$

Note that

$$d^o_{\cosh}(f,g) = 2\left(e^{d_I^*(f,g)} - 1\right) \tag{4.23}$$

9) Symmetrized Model Distortion

$$d_{cm}^{(q)}(f,g) = \left\{ \left\| 1 - \frac{f^+}{g^+} \right\|_2^q + \left\| 1 - \frac{g^+}{f^+} \right\|_2^q \right\}_1^{1/q} \quad , \tag{4.24}$$

$$d_{ncm}^{(q)}(f,g) = d_{cm}^{(q)}(f/\sigma_f^2, g/\sigma_g^2) \quad , \tag{4.25}$$

$$d_{nm}^{(q)}(f,g) = \left\{ \left\| 1 - \frac{f^{1/2}}{g^{1/2}} \right\|_2^q + \left\| 1 - \frac{g^{1/2}}{f^{1/2}} \right\|_2^q \right\}^{1/q} \quad . \tag{4.26}$$

10) Symmetrized L_1 Ratio Distortion

$$d_1^{(1)}(f,g) = \|1 - f/g\|_1 + \|1 - g/f\|_1 \tag{4.27}$$

This has another form. Using the fact that

$$|1-a| + |1-1/a| = |a - 1/a| \quad , \quad a \geq 0 \quad , \tag{4.28}$$

we have

$$d_1^{(1)}(f,g) = \|f/g - g/f\|_1 \quad , \tag{4.29}$$

which is an L_1 analog to the cosh measure.

Many other distortion measures can be defined by combining the previous measures as in Section 3, but the preceeding are the basic measures considered here.

5. PROPERTIES AND INTERRELATIONS OF DISTORTION MEASURES

The Innovations Property

Many of the distortion measures considered have the interesting property that the distortion between f and g is bound below by the distortion between two white processes having the same gains, that is, the innovations processes of f and g. We say that a distortion measure has the innovations property if $d(f,g) \geq d(\sigma_f^2,\sigma_g^2)$. Clearly this is a trivial property for gain-normalized distortion measures since then $d(\sigma_f^2,\sigma_g^2) = d(1,1) = 0 \leq d(f,g)$. We have that since $r_{f/g}(0) \geq \sigma_f^2/\sigma_g^2$,

$$d_{IS}(f,g) \geq d_{IS}(\sigma_f^2,\sigma_g^2) \quad . \tag{5.1}$$

Thus

$$\begin{aligned} d_{cosh}(f,g) &= \frac{1}{2}\,(d_{IS}(f,g) + d_{IS}(g,f)) \\ &\geq \frac{1}{2}\,(d_{IS}(\sigma_f^2,\sigma_g^2) + d_{IS}(\sigma_g^2,\sigma_f^2)) \\ &= d_{cosh}(\sigma_f^2,\sigma_g^2) = \frac{1}{2}\left|\frac{\sigma_f}{\sigma_g} - \frac{\sigma_g}{\sigma_f}\right|^2 \quad . \end{aligned} \tag{5.2}$$

Equation (5.2) shows that if $d_{cosh}(f,g_n) \to 0$ as $n \to \infty$, then necessarily $\sigma_{g_n}^2 \to \sigma_f^2$. From (4.12)

$$\begin{aligned} d_{cm}(f,g)^2 &= 1 + r_{f/g}(0) - 2\sigma_f/\sigma_g \geqq 1 + \sigma_f^2/\sigma_g^2 - 2\sigma_f/\sigma_g \\ &= |1 - \sigma_f/\sigma_g|^2 \\ &= d_{cm}(\sigma_f^2,\ \sigma_g^2)^2 \end{aligned}$$

whence

$$d_{cm}(f,g) \geq d_{cm}(\sigma_f^2,\sigma_g^2) \quad , \tag{5.3}$$

and

$$d_{cm}^{(1)}(f,g) \geq d_{cm}^{(1)}(\sigma_f^2,\sigma_g^2) \quad . \qquad (5.4)$$

We have

$$d_{log}(f,g) = \|\ell n\ f/g\|_p \geq \|\ell n\ f/g\|_1$$

$$= (2\pi)^{-1}\int_{-\pi}^{\pi} |\ell n\ f(\Theta)/g(\Theta)|d\Theta \geq \left|(2\pi)^{-1}\int_{-\pi}^{\pi} \ell n\ f(\Theta)/g(\Theta)d\Theta\right|$$

$$= |\ell n\ \sigma_f^2/\sigma_g^2| = \|\ell n\ \sigma_f^2/\sigma_g^2\|_p$$

and therefore

$$d_{log}(f,g) \geq d_{log}(\sigma_f^2,\sigma_g^2) \quad . \qquad (5.5)$$

It is not known if the L_1 ratio or noncausal model distortions possess this property.

Nonsymmetric Distortion Implications

We have that

$$(2\pi)^{-1}\int_{-\pi}^{\pi} \frac{f(\Theta)^{1/2}}{g(\Theta)^{1/2}}d\Theta \geq \sigma_f/\sigma_g \qquad (5.6)$$

and hence from (4.12)-(4.13)

$$d_{cm}(f,g) \geq d_{nm}(f,g) \qquad (5.7)$$

so that $d_{cm} \Rightarrow d_{nm}$. We have from (4.4) and (4.12) that

$$d_{IS}(f,g) = d_{cm}(f,g)^2 - \left|1 - \frac{\sigma_f}{\sigma_g}\right|^2 + \left(\frac{\sigma_f^2}{\sigma_g^2} - 1 - \ell n\ \frac{\sigma_f^2}{\sigma_g^2}\right)$$

$$= d_{cm}(f,g)^2 - d_{cm}(\sigma_f^2,\sigma_g^2)^2 + d_{IS}(\sigma_f^2,\sigma_g^2) \qquad (5.8)$$

and hence

$$d_{IS}(f,g) - d_{IS}(\sigma_f^2,\sigma_g^2) = d_{cm}(f,g)^2 - d_{cm}(\sigma_f^2,\sigma_g^2)^2$$

so that additively removing the innovations distortion makes the Itakura-Saito and causal model distortions the same. Expanding and cancelling in (5.8) yields

$$d_{IS}(f,g) = d_{cm}(f,g)^2 + 2\ \frac{\sigma_f}{\sigma_g} - 1 - \ell n\ \frac{\sigma_f}{\sigma_g} \geq d_{cm}(f,g)^2 \tag{5.9}$$

and hence $d_{IS}(f,g) \Rightarrow d_{cm}(f,g) \Rightarrow d_{nm}(f,g)$. Next observe that if $x \to 1$, then $x-1-\ell n\ x \to 0$. Thus if $d_{cm}(f,g_n) \to 0$ as $n \to \infty$, from the innovations property $|1-\sigma_f/\sigma_{g_n}| \underset{n\to\infty}{\to} 0$ whence $(\sigma_f/\sigma_{g_n} -1-\ell n(\sigma_f/\sigma_{g_n})) \to 0$ and therefore $d_{cm}(f,g_n) \to 0$ implies $d_{IS}(f,g_n) \to 0$. We have thus shown that

$$d_{IS}(f,g) \Leftrightarrow d_{cm}(f,g) \Rightarrow d_{nm}(f,g) \tag{5.10}$$

We have from (4.11) that $d_{ncm}(f,g) \Rightarrow d_I(f,g)$ and from (4.10) and (4.12) that

$$d_{ncm}(f,g)^2 = \frac{\sigma_g^2}{\sigma_f^2}\ \{d_{cm}(f,g)^2 - d_{cm}(\sigma_f^2,\sigma_g^2)^2\} \leq \frac{\sigma_g^2}{\sigma_f^2}\ d_{cm}(f,g)^2$$

and therefore using the innovations property of d_{cm}

$$d_{ncm}(f,g) \leq \frac{\sigma_g}{\sigma_f}\ d_{cm}(f,g) \leq \frac{d_{cm}(f,g)}{1-d_{cm}(f,g)} \tag{5.11}$$

which implies

$$d_{cm} \Rightarrow d_{ncm} \tag{5.12}$$

Recall from (4.10) that

$$d_{ncm}(f,g)^2 = e^{d_I(f,g)} - 1$$

which implies that

$$d_{ncm} \Rightarrow d_I \tag{5.13}$$

since

$$d_I(f,g) = \ell n(d_{ncm}(f,g)^2 + 1) \le d_{ncm}(f,g)^2 \tag{5.14}$$

and if $d_I(f,g_n) \to 0$ as $n \to \infty$, then

$$d_{ncm}(f,g_n) = \sum_{k=1}^{\infty} \frac{d_I(f,g_n)^k}{k!} \underset{n\to\infty}{\to} 0 \tag{5.15}$$

From the inequality

$$x^{1/2} \ge \min(1,x) \quad , \quad x \ge 0 \tag{5.18}$$

we have from (4.13) and (4.16) that

$$d_{nm}(f,g)^2 = \|1-f^{½}/g^{½}\|_2^2 \le \|1-f/g\|_1 \tag{5.19}$$

and therefore

$$d_1(f,g) \Rightarrow d_{nm}(f,g) \quad , \tag{5.20}$$

Note that (5.19) also follows from the definitions and the inequality

$$|1-x| \ge |1-x^{½}|^2 \tag{5.21}$$

(which follows from (5.18)).

The implications for nonsymmetric distortion measures are summarized below.

$$\begin{array}{ccc} d_I & \not\Leftarrow\!\!\Rightarrow & d_{ncm} \\ & \Uparrow & \\ d_{IS} & \Leftrightarrow & d_{cm} \\ & \Downarrow & \\ & d_{nm} \Leftarrow d_1 & \end{array} \tag{5.22}$$

Note that d_{ncm} can be written in a manner strongly resembling d_1 as follows:

$$d_{ncm}(f,g) = \sigma_g^2 r_{f/g}(0)/\sigma_f^2 - 1 = \left\| \frac{f/\sigma_f^2}{g/\sigma_g^2} - 1 \right\|_1 ,$$

that is, $d_{ncm}(f,g)$ is a gain-normalized version of $d_1(f,g)$, i.e.,

$$d_{ncm}(f,g) = d_1(f/\sigma_f^2, g/\sigma_g^2)$$

$$= d_{cm}(f/\sigma_f^2, g/\sigma_g^2) \qquad (5.23)$$

This does not imply directly that $d_1 \Rightarrow d_{ncm}$ since we have been unable to show that $d_1 \to 0$ implies that $\sigma_f^2/\sigma_g^2 \to 1$. It can be easily shown, however, that

$$\|1-f/g\|_1 + |1-\sigma_f^2/\sigma_g^2| = d_1(f,g) + d_1(\sigma_f^2,\sigma_g^2) \Rightarrow d_{ncm} \quad . \qquad (5.24)$$

Finally, if we consider coding equivalence, since d_I and d_{ncm} are related monotonically by (4.10), clearly $d_{ncm} \subset\supset d_I$ and from the stated properties $d_{IS} \supset d_I$. Thus

$$\begin{array}{ccc} & d_{IS} & \\ & \cup & \\ d_{ncm} & \subset\supset & d_I \end{array} \qquad (5.25)$$

Symmetric Distortion Implications

We first focus on the cosh distortion and log spectral deviation as these are the most commonly proposed measures in the speech literature. We then develop their relations to the other measures. Gray and Markel [2] proved graphically that

$$\|\ln f/g\|_2^2 \le 2d_{\cosh}(f,g)$$

and hence that $d_{\cosh} \Rightarrow d_{\log}$. Implicit in their proof is the following Taylor series expansion argument: For real $x \ge 0$ set $x = e^{\alpha}$ and we have

$$\begin{aligned}
|x^{1/2}-x^{-1/2}| &= |e^{\alpha/2}-e^{-(\alpha/2)}| \\
&= \left|\sum_{k=0}^{\infty} \frac{(\alpha/2)^k}{k!} - \sum_{k=0}^{\infty} \frac{(-\alpha/2)^k}{k!}\right| = 2\left|\sum_{k=1,3,5,\ldots} \frac{(\alpha/2)^k}{k!}\right| \\
&= \left|\alpha + 2\sum_{k=3,5,\ldots} \frac{(\alpha/2)^k}{k!}\right| = |\alpha| + 2\left|\sum_{k=3,5,\ldots} \frac{(\alpha/2)^k}{k!}\right| \ge |\alpha| \\
&= |\ln x|
\end{aligned}$$

whence

$$\|\ln f/g\|_2^2 \le \|f^{1/2}/g^{1/2} - g^{1/2}/f^{1/2}\|_2^2 = 2d_{\cosh}(f,g) \tag{5.26}$$

The converse implication is not in general true as can easily be seen by counterexample.

From (4.12), (4.19), and (5.7) we have that

$$2d_{\cosh}(f,g)^2 = r_{f/g}(0) + r_{g/f}(0) - 2$$

$$= d_{cm}(f,g)^2 + d_{cm}(g,f)^2 + 2\sigma_f/\sigma_g + 2\sigma_g/\sigma_f - 4$$

$$= d_{cm}^{(2)}(f,g)^2 + 2(\sigma_f^{1/2}/\sigma_g^{1/2} - \sigma_g^{1/2}/\sigma_f^{1/2})^2$$

$$\geq d_{cm}^{(2)}(f,g)^2 \geq d_{nm}^{(2)}(f,g)^2 \tag{5.27}$$

and therefore $d_{\cosh} \Rightarrow d_{cm}^{(2)} \Rightarrow d_{cm}^{(2)}$. From (5.7)

$$d_{cm}^{(2)}(f,g) \geq d_{nm}^{(2)}(f,g) \quad . \tag{5.28}$$

We also have that from (3.3)

$$(2d_{\cosh}(f,g))^{1/2} = \|f^{1/2}/g^{1/2} - g^{1/2}/f^{1/2}\|_2 \leq \|1-f^{1/2}/g^{1/2}\|_2 + \|1-g^{1/2}/f^{1/2}\|_2$$

$$= d_{nm}^{(1)}(f,g) \leq 2^{1/2} d_{nm}^{(2)}(f,g)$$

so that

$$d_{\cosh}(f,g) \leq d_{nm}^{(2)}(f,g)^2 \quad . \tag{5.29}$$

To summarize we have that

$$\frac{1}{2} d_{nm}^{(2)}(f,g)^2 \leq \frac{1}{2} d_{cm}^{(2)}(f,g)^2 \leq d_{\cosh}(f,g) \leq d_{nm}^{(2)}(f,g)^2 \leq d_{cm}^{(2)}(f,g)^2 \tag{5.30}$$

which proves that

$$d_{\cosh} \Leftrightarrow d_{cm}^{(q)} \Leftrightarrow d_{nm}^{(q)} \quad , \quad q=1,2,\ldots \tag{5.31}$$

Analagous to (4.10) and (5.13)-(5.15) we have from (4.23) that

$$\begin{gathered} d_{\cosh}^{o}(f,g) \Leftrightarrow d_I^*(f,g) \\ d_{\cosh}^{o}(f,g) \subset\supset d_I^*(f,g) \\ d_I^*(f,g) \leq \frac{1}{2} d_{\cosh}^{o}(f,g) \quad . \end{gathered} \tag{5.32}$$

We also have the following: $d_I \Leftrightarrow d_{ncm}$ and therefore

$$d_I^* \Leftrightarrow d_{ncm}^{(q)} \quad , \tag{5.33}$$

$d_{cm} \Rightarrow d_{nm}$, and therefore

$$d_{cm}^{(q)} \Rightarrow d_{nm}^{(q)} \quad , \tag{5.34}$$

$d_1 \Rightarrow d_{nm}$, and therefore

$$d_1^{(q)} \Rightarrow d_{nm}^{(q)} \quad . \tag{5.35}$$

For example,

$$d_1^{(1)}(f,g) = \|f/g - g/f\|_1 \geq 2d_{cosh}(f,g)^2 \tag{5.30}$$

To summarize for the symmetric case

$$\begin{array}{ccccccc}
 & & & d_{log} & & & \\
 & & & \Uparrow & & & \\
d_{cosh} & \Leftrightarrow & d_{cm}^{(q)} & \Leftrightarrow & d_{nm}^{(q)} & \Leftarrow & d_1^{(q)} \\
 & & & \Downarrow & & & \\
 & & d_{cosh}^{o} & \Leftrightarrow & d_I^* & \Rightarrow & d_{ncm}^{(q)}
\end{array} \tag{5.37}$$

and

$$d_{cosh}^{o} \subset\supset d_I^* \subset\supset d_{ncm}^{(2)} \tag{5.38}$$

since all are minimized by minimizing $\sigma_g^2 r_{f/g}(0)$.

Coupled with the fact that all symmetrized distortions imply their unsymmetrized versions, this completes the catalogue of equivalence and implication relations.

6. SOME TIME DOMAIN EXPRESSIONS

The emphasis here has been on spectral representations. For completeness and to ease comparison of the distortion measures defined here with alternate forms appearing in the literature we observe some time-domain expressions.

Most of the distortion measures involve the term $r_{f/g}(0)$. Transforming this equation yields

$$r_{f/g}(0) = (2\pi)^{-1} \int_{-\pi}^{\pi} \frac{f(\Theta)}{g(\Theta)} \, d\Theta$$

$$= \sum_{k=0}^{\infty} r_f(k) r_{1/g}(k)$$

and hence if $g = \sigma_g^2/|A|^2$ we have

$$r_{f/g}(0) = \frac{1}{\sigma_g^2} \sum_{k=0}^{\infty} r_f(k) r_{|A|^2}(k) \quad .$$

Alternatively, if we think of this as the power in the output of a filter $1/g^+ = A/\sigma_g$ with an input process $\{X_n\}$ with spectral density f, then

$$r_{f/g}(0) = \frac{1}{\sigma_g^2} E \left\{ \left| \sum_{k=0}^{\infty} a_k X_{n-k} \right|^2 \right\}$$

$$= \frac{1}{\sigma_g^2} \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} a_k a_j r_f(k-j)$$

where the sum will be finite if A has finite order. This can also be expressed in matrix notation

$$r_{f/g}(0) = \frac{1}{\sigma_g^2} \underset{\sim}{a}' \underset{\sim}{R} \, \underset{\sim}{a}$$

where $\underset{\sim}{a}' = (1, a_1, a_2, \ldots)$ is a semi-infinite vector and $\underset{\sim}{R}$ the doubly

semi-infinite correlation matrix. If A has finite order m this reduces to

$$r_{f/g}(0) = \frac{1}{\sigma_g^2}(a^m)' T_m(f) a^m$$

where $(a^m)' = (1, a_1, \ldots, a_{m-1})$. In addition, the theory of Toeplitz forms [8,16,17] can be used to write

$$\sigma_g^2 r_{f/g}(0) = \lim_{n \to \infty} n^{-1} \mathrm{tr} T_n(f) T_n(|A|^2) \quad .$$

Lastly we observe that if $f = \sigma_f^2/|B|^2$, then

$$d_{ncm}(f,g)^2 = \left\| 1 - \frac{f^+/\sigma_f}{g^+/\sigma_g} \right\|_2^2$$

$$= \|1 - A/B\|_2^2 = \frac{1}{\sigma_f^2}(2\pi)^{-1} \int_{-\pi}^{\pi} \left| \sum_{k=0}^{\infty} (b_k - d_k) e^{-ik\Theta} \right|^2 f(\Theta) d\Theta$$

$$= \frac{\sum_{k=0}^{\infty} \sum_{j=0}^{\infty} (b_k - d_k) r_f(k-j)(b_j - a_j)}{\sum_{k=0}^{\infty} b_k r_f(k-j) b_j}$$

$$= (\underset{\sim}{b} - \underset{\sim}{a})' \underset{\sim}{R} (\underset{\sim}{b} - \underset{\sim}{a}) / \underset{\sim}{b}' \underset{\sim}{R}\, \underset{\sim}{b} \quad ,$$

the form used by Itakura [18] and Chaffee [19].

7. IMPLICATIONS FOR SPEECH COMPRESSION

In this section some implications and applications of the preceeding results to speech compression are discussed and some research directions proposed. The mathematical model adopted herein is the following. We are given a sequence of "symbols" $\{f_n\}_{n=-\infty}^{\infty}$ where each symbol $f_n \in \mathcal{F}$ is itself the power spectral density of a stationary and ergodic nondeterministic random process as in Section 2, that is, the alphabet $\mathcal{F}$ consists of all nonnegative even real valued functions $f(\Theta)$, $\Theta \in [-\pi,\pi]$, for which $f, f^{-1} \in L_1$ (and hence $\ln f \in L_1$). In actual practice each f_n is obtained via a transformation on a windowed speech waveform $\{x(t); t \in [nT,(n+1)T)\}$ by, for example, forming the magnitude square of the discrete Fourier transform. We are also given a finite reproduction alphabet $\hat{\mathcal{F}} \subseteq \mathcal{F}$. A data compression system maps each $f_n \in \mathcal{F}$ into a reproduction $\hat{f}_n \in \hat{\mathcal{F}}$ and then a binary index (fixed or variable length) is transmitted to a receiver who then reconstructs $\hat{f}_n$. The goal is to minimize the average distortion $d(f_n,\hat{f}_n)$ for a fixed code rate. We note that the theory of source coding is valid for such a general alphabet and distortion measure provided that the source $\{f_n\}$ and reproduction set $\hat{\mathcal{F}}$ are such that $d(f_n,\hat{f}_n) = \infty$ cannot occur with nonzero probability. This is a physical assumption without which finite average distortion communication is not possible.

A general data compression system maps several input symbols into several output symbols: $(f_{nN},\ldots,f_{n(N+1)-1}) \rightarrow (\hat{f}_{nN},\ldots,\hat{f}_{n(N+1)-1})$, $n=0, \pm 1, \pm 2,\ldots$ a technique called block coding; or several input symbols into a single output symbol: $(f_{n-N},\ldots,f_n,\ldots,f_{n+N}) \rightarrow \hat{f}_n$,

$n=0,\pm 1,\ldots,$ a technique called sliding-block coding [21,22]. We here consider the special case N=1 of "single-symbol quantization" of $\mathcal{N}$, a mapping of the form $f_n \to \hat{f}_n$. LPC compression systems using the autocorrelation method have this form where a "two-step" compressor is used. The first step is to transform f into another spectrum $\tilde{f}$ that is the best m^{th} order autoregressive model of f in the sense that $\tilde{f} = \tilde{f}(f) \in \mathcal{N}_m$ = {all spectral densities of the form $\sigma^2/|A(e^{i\theta})|^2 \in \mathcal{N}$, where $A(z) = 1+\sum_{k=1}^{m} a_k z^{-k}$} and "best" means $\tilde{f}(f)$ minimizes $d_{IS}(f,\tilde{f})$. This is a "system identification" step and results in distortion [23]

$$d_{IS}(f,\tilde{f}(f)) = \ln(\sigma_m^2/\sigma_f^2) \tag{7.1}$$

In the next step $\tilde{f} \to \hat{f} \in \hat{\mathcal{N}} \subset \mathcal{N}_m \subset \mathcal{N}$, where $\hat{\mathcal{N}}$ is a finite collection of m^{th} order autoregressive models. There are several ways to construct $\hat{\mathcal{N}}$, the most common being to transform the model $\tilde{f}$ into a vector of reflection coefficients and gain and separately quantize each according to some criterion [5 , 6]. Several criterion are possible for this real number quantization, but theory and practice have shown that most sensible approaches yield nearly equivalent results [6]. The rate of such a system is $\log_2|\hat{\mathcal{N}}|$ bits per "symbol," where here symbol means a windowed speech waveform of typically 20ms.

Such a system is ad hoc and nonoptimal since, for example, "optimal" usually means minimizing distortion for a fixed $\hat{\mathcal{N}}$, yet here one distortion measure (d_{IS}) is used in the first step and another (often a magnitude error on reflection coefficients which is approximately equal to $d_{\log}$ on the spectra) in the second. An optimal compressor

according to d_{IS}, for example, would take f_n and directly find the $\hat{f} \in \hat{\mathcal{M}}$ minimizing $d_{IS}(f,\hat{f})$ to form $\hat{f}_n(f)$ (no other algorithm can yield a lower Itakura-Saito distortion overall). An alternative conceptually optimum system would be to use a two-step compressor as before, with the second step forming an optimal quantizer of the $\tilde{f} \in \mathcal{M}_m$, that is, form $\tilde{f}(f_n)$ as before and then set $\hat{f}(\tilde{f}) \in \hat{\mathcal{M}}$ as the model minimizing $d_{IS}(\hat{f},\tilde{f})$. With this system, however, an immediate problem arises. Since d_{IS} is not metric, how do we know that a good job in each step ($d_{IS}(f_n,\tilde{f}(f_n))$ and $d_{IS}(\tilde{f}(f_n), \hat{f}(\tilde{f}(f_n)))$ small) will yield a good job overall ($d_{IS}(f_n,\hat{f}(\tilde{f}(f_n)))$ small)? This leads us to one of the interesting properties of the Itakura-Saito distortion -- it does have a triangle inequality (with equality, actually) in two-step systems with the first step as above and the two LPC systems described yield the same encoding. To see this let $f^{(m)} \in \mathcal{M}_m$ minimize $d_{IS}(f,\tilde{f})$ over $\tilde{f} \in \mathcal{M}_m$ (and hence as previously $f^{(m)}(\Theta) = \sigma_m^2/|\hat{A}(e^{i\Theta})|^2$ where $\hat{A}$ satisfies (2.10)), let $\hat{f} \in \hat{\mathcal{M}} \subseteq \mathcal{M}_m$, and use (2.14) and (7.1) to write

$$
\begin{aligned}
d_{IS}(f,\hat{f}) &= (2\pi)^{-1} \int_{-\pi}^{\pi} \frac{f(\Theta)}{\hat{f}(\Theta)}\, d\Theta - 1 - \ln \sigma_f^2/\sigma_{\hat{f}}^2 \\
&= (2\pi)^{-1} \int_{-\pi}^{\pi} \frac{f^{(m)}(\Theta)}{\hat{f}(\Theta)}\, d\Theta - 1 - \ln \sigma_f^2/\sigma_{\hat{f}}^2 \\
&= (2\pi)^{-1} \int_{-\pi}^{\pi} \frac{f^{(m)}(\Theta)}{\hat{f}(\Theta)}\, d\Theta - 1 - \ln \sigma_m^2/\sigma_{\hat{f}}^2 + \ln(\sigma_m^2/\sigma_f^2) \\
&= d_{IS}(f,f^{(m)}) + d_{IS}(f^{(m)},\hat{f}) \qquad . \qquad (7.2)
\end{aligned}
$$

Thus for this type of two step system, regardless of the quantization

rule, the overall Itakura-Saito distortion is exactly the sum of the incurred distortion in the separate system-identification and compression steps. In particular, if the compression of $f^{(m)}$ is done optimally for a fixed set $\hat{\mathcal{T}}$, then it is equivalent to an optimum quantizer operating directly on f! It would be of interest to compute d_{IS} for real overall systems since this would yield a distortion measure consistent with the implicit definition of optimum in the first step. It would also be of interest to see which of the various reflection coefficient quantization rules yield the smallest quantization (and hence overall) d_{IS}. We conjecture that as in [6], the results would be very nearly the same since little improvement is possible when one is constrained to separately quantize each real parameter. It would also be interest to see if computationally efficient approximately optimal (in d_{IS}) mappings $\mathcal{T}_m \to \hat{\mathcal{T}}$ could be developed.

Next suppose that subjective testing might indicate that some other distortion measure is better than d_{IS}. This alternative distortion could be used in either of the two kinds of systems -- the two-step or direct quantization. In a two-step system the use of an alternative and symmetric measure such as the cosh or log spectral deviation would result in finding a model that matched zeroes as well as poles and hence $\mathcal{T}_m$ might be replaced by a collection of mixed moving-average and autoregressive (ARMA) models. Unfortunately, however, finding a best finite-order model matching poles and zeroes (or even zeroes alone) seems a very difficult problem. This points out that one of the nice facets of d_{IS} is the simplicity and speed with which the minimum distortion $f^{(m)} \in \mathcal{T}_m$ can be found. This suggests compromises: If an

alternative distortion measure is suggested by subjective testing, but scatter plots such as those of Markel and Gray [2] or Matsuyama [24] suggest that for small distortion the alternative measure is highly correlated with d_{IS}, then use d_{IS} as an approximation in Step 1 to facilitate computation and then use the other measure in Step 2. Performance can be slightly improved by replacing σ_m^2 in Step 1 by the gain-optimized value for $\hat{A}$ according to the other distortion measure, e.g., σ_g^2 for the cosh measure.

If a two step system is constructed using the causal filter measure, then the behavior is similar to that of the Itakura-Saito system and one again obtains a triangle inequality. For $g = \sigma^2/|A|^2 \in \mathcal{M}_m$ we have that

$$d_{cm}(f,\sigma^2/|A|^2)^2 = 1+\sigma^2 r_{f|A|^2}(0) - 2\sigma_f/\sigma$$

which is minimized by minimizing $r_{f|A|^2}(0)$ by choosing A to satisfy (2.10) and by choosing the optimum gain (see (4.13)) $\sigma = \sigma_m^2/\sigma_f \geq \sigma_m$. Thus, the monic filter is the same as that of the Itakura-Saito distance, but the gain is larger. Denote the resulting spectrum $\tilde{f}(\theta) = \tilde{\sigma}/|A(e^{i\theta})|^2$ where

$$\tilde{\sigma} = \sigma_m^2/\sigma_f$$

$$d_{cm}(f,\tilde{f})^2 = \min_{g \in \mathcal{M}_m} d_{cm}(f,g)^2$$

$$= 1 - (\sigma_f/\sigma_m)^2$$

Let $\hat{f} = \hat{\sigma}^2/|\hat{A}|^2$ denote the quantized version of $\tilde{f}$ resulting from the second step and we have from (2.14) that

$$d_{cm}(f,\hat{f})^2 = 1 + (2\pi)^{-1} \int_{-\pi}^{\pi} \frac{f(\theta)}{\hat{f}(\theta)}\, d\theta - 2\sigma_f/\hat{\sigma}$$

$$= 1 + (2\pi)^{-1} \int_{-\pi}^{\pi} \frac{|\hat{A}(e^{i\theta})|^2}{\hat{\sigma}^2} \frac{\sigma_m^2}{|A(e^{i\theta})|^2}\, d\theta - 2\sigma_f/\hat{\sigma}$$

$$= 1 + \frac{\sigma_m^2}{\tilde{\sigma}^2} (2\pi)^{-1} \int_{-\pi}^{\pi} \frac{\tilde{f}(\theta)}{\hat{f}(\theta)}\, d\theta - 2\sigma_f/\hat{\sigma}$$

$$= \frac{\sigma_m^2}{\tilde{\sigma}^2} d_{cm}(\tilde{f},\hat{f})^2 - \frac{\sigma_m^2}{\tilde{\sigma}^2} + 2\frac{\sigma_m^2}{\tilde{\sigma}^2} \cdot \frac{\tilde{\sigma}}{\hat{\sigma}} - 2\sigma_f/\hat{\sigma}$$

$$= \frac{\sigma_f^2}{\sigma_m^2} d_{cm}(\tilde{f},\hat{f})^2 - \frac{\sigma_f^2}{\sigma_m^2} = \frac{\sigma_f^2}{\sigma_m^2} d_{cm}(\tilde{f},\hat{f})^2 + d_{cm}(\tilde{f},f)^2$$

$$\leq d_{cm}(\tilde{f},\hat{f})^2 + d_{cm}(f,\hat{f})^2 ,$$

a sort of triangle inequality.

Another observation on these systems is that implicit (and relevant) subjective testing can be accomplished for the various distortion measures by simulating either type of compression system for a given distortion (likely using d_{IS} in Step 1 of the two-step for simplicity) and then listening to the reconstructed output. A reproduction set could be taken as the LPC system reproduction set or, say, the direct quantization reproduction set of Chaffee [19], who uses d_{ncm} to select the monic filter part of $\hat{f}$ and alternative criteria to select the gain and pitch. The point is, a good subjective test of a distortion measure is to listen to the output of a minimum distortion compressor using that measure. We are currently attempting to study various rules for selecting the finite reproduction set $\hat{\mathcal{R}}$ from observed data and for efficiently computing the various $d(f,\hat{f})$ and finding the best $\hat{f}$. If such systems are successful (in particular, if they are comparable to LPC systems as Chaffee's [19] work suggests), our feeling is they will provide an alternative to LPC requiring less

computation but more memory. Success in this approach would also open two other avenues of future research: (1) Compression systems using block or sliding-block coding on the $\{f_n\}$ could be attempted. This may sound prohibitively complex, but if the single symbol systems were well-understood, then the "Fake Process" approach to data compression [25,26] would provide a straightforward ad hoc technique for improving performance using the memory in the $\{f_n\}$. (2) If a large (high rate) finite class $\hat{\eta}$ could be shown to be "rich" enough to well-model most $\{f_n\}$, then the long run probabilistic behavior could be approximated by compiling first-order histograms of occurrences of $\hat{f}$ in real speech. The probabilistic model could be coupled with the distortion measure and the Blahut algorithm [27] to compute a meaningful distortion rate function for speech and thereby obtain an absolute unbeatable bound on performance of single-symbol direct quantizers (such as the LPC and Chaffee systems). It would be interesting to know how nearly "optimal" the ad hoc but highly successful LPC systems are and whether or not one must resort to block or sliding-block coding in order to obtain real improvement over LPC systems. If efficient means of finding conditional histograms could be found, higher order distortion rate functions could be obtained yielding performance bounds on more general system.

Another problem has to be addressed in simulating such systems, that of decoding. How does one convert a spectral density $\hat{f}$ back into a sound? Mathematically, there exists a random process having such a density. In fact there exist many (Gaussian, for example), and it is important to know which to use. Again mathematically, it makes no difference insofar as the distortion measures herein considered are

concerned since these all place zero distortion on two identical spectra. Practically, however, these distortion measures simply approximate the biological distortion measure of the human brain and the actual process used in reconstruction will very likely make a difference in subjective quality. Here we can only propose to try ad hoc techniques when this stage is reached. A first try would be to rise simply Gaussian noise driving the factored spectrum to produce a Gaussian process with the correct (optimal) spectral density. This does not mean that speech "looks" Gaussian, only that Gaussian pseudo-speech may sound like speech. Alternatively, some work [28,29] indicates a double-sided exponentially distributed white process may perform more satisfactorily. This is a problem which must be treated experimentally with human listeners as the distortion measures cannot tell the difference of underlying statistics except through the spectra.

This report was motivated by the research described in this section. It was found useful to have a catalogue of the various distortion measures, properties, and interrelations. As the experiments proposed here will likely involve considerable time to reach any solid conclusions, the preliminary work on the distortion measures has been collected now in the hopes of being useful to others conducting similar research.

REFERENCES

[1] D.R. Brillinger, Time Series, Holt, Rinehart and Winston Inc., N.Y., 1975.

[2] A.H. Gray, Jr. and J.D. Markel, "Distance Measures for Speech Processing," IEEE Trans.-ASSP, Vol. 24, pp. 380-391, Oct. 1976.

[3] R. Viswanathan, J. Makhoul, and W. Russel, "Towards Perceptually Consistent Measures of Spectral Distance," 1976 Internat'l. Conference on Acoustics, Speech & Signal Processing, Philadelphia, PA, April 1976.

[4] J. Makhoul, "Spectral Analysis of Speech Linear Prediction," IEEE Trans. AU, Vol. 21, pp. 140-148, 1973.

[5] R. Viswanathan and J. Makhoul, "Quantization Properties of Transmission Parameters in Linear Predictive Systems," IEEE Trans.-ASSP, Vol. 23, pp. 309-321, 1975.

[6] A.H. Gray, Jr., R.M. Gray, and J.D. Markel, "Comparison of Optimal Quantizations of Speech Reflection Coefficients," IEEE Trans.-ASSP, Vol. 25, pp. 9-23, Feb. 1977.

[7] R.B. Ash, Real Analysis and Probability, Academic Press, N.Y., 1972.

[8] U. Grenander and G. Szegö, Toeplitz Forms and Their Applications, Univ. of Calif. Press, Berkeley, Calif., 1958.

[9] J.L. Doob, Stochastic Processes, John Wiley & Sons, New York, 1953.

[10] L.Ya. Geronimus, Orthogonal Polynomials, Consultants Bureau, Noble Offset Printers, Inc., N.Y., 1961.

[11] T. Berger, Rate Distortion Theory, Prentice Hall, Englewood Cliffs, N.J., 1971.

[12] F. Itakura and S. Saito, "Analysis Synthesis Telephony Based Upon the Maximum Likelihood Method," Reports of the 6th Internat'l. Cong. Acoust., Y. Kohasi, ed., Tokyo, C-5-5, C17-20, 1968.

[13] S. Kullback, Information Theory and Statistics, Dover Press, N.Y., 1968.

[14] T. Kailath, "The Divergence and Bhattacharyya Distance Measures in Signal Selection," IEEE Trans. Com. Tech., Vol. 15, pp. 52-60, 1967.

[15] M. Pinsker, Information and Information Stability of Random Variables and Processes, Moscow: Ize. Akad. Nauk. SSSR, 1960 (Translation: Holden-Day, San Francisco, Calif., 1964).

[16] R.M. Gray, "On the Asymptotic Eigenvalue Distribution of Toeplitz Matrices," IEEE Trans.-IT, Vol. 18, pp. 725-730, Nov. 1972.

[17] R.M. Gray, Toeplitz and Circulant Matrices: II, Stanford University Information Systems Laboratory Technical Rept. No. 6504-1, April, 1977.

[18] F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," IEEE Trans.-ASSP, Vol. 23, pp. 67-72, Feb. 1975.

[19] D.L. Chaffe, "Applications of Rate Distortion Theory to the Bandwidth Compression of Speech Signals," Ph.D. Dissertation, Univ. of Calif. at Los Angeles, 1975.

[20] R.M. Gray, D.L. Neuhoff and P.C. Shields, "A Generalization of Ornstein's $\bar{d}$- Distance with Applications to Information Theory," Ann. Prob., Vol. 3, No. 2, pp. 315-328, 1975.

[21] L.D. Davisson and R.M. Gray, eds., Data Compression, Dowden, Hutchinson and Ross, Inc., Stroudsburg, PA, 1976.

[22] R.M. Gray and L.D. Davisson, eds., Ergodic and Information Theory, Dowden, Hutchinson, and Ross, Stroudsburg, PA, 1977.

[23] J.D. Markel and A.H. Gray, Jr., Linear Prediction of Speech, Springer-Verlag, N.Y., 1976.

[24] Y. Matsuyama, Ph.D. Dissertation, Stanford University, 1978.

[25] Y. Linde and R.M. Gray, "A Fake Process Approach to Data Compression," IEEE Trans.-COM., Vol. 26, pp. , June, 1978.

[26] Y. Linde and R.M. Gray, The Design of Tree and Trellis Data Compression Systems, Stanford University Information Systems Lab. Technical Report No. 6504-2, Feb. 1978.

[27] R.E. Blahut, "Computation of Channel Capacity and Rate-Distortion Functions," IEEE Trans.-IT, Vol. 18, pp. 460-473, July, 1972.

[28] L.H. Zettenberg and J. Uddenfeld, "Adaptive Delta Modulation with Delayed Decision," IEEE Trans. Comm., Vol. 25, pp. 408-417, 1977.

[29] S. Tazaki, H. Osawa and Y. Yameda, "Generation of Pseudo-Voice Signal Based on a Simple Algorithm and Its Characteristics," Trans. IECE, Vol. 59, pp. 776-777 (in Japanese), Sept. 1976.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 6504-2	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Spectral Distortion Measures for Speech Compression		5. TYPE OF REPORT & PERIOD COVERED Technical Report 5-1-73 to 4-30-78
		6. PERFORMING ORG. REPORT NUMBER SEL-78-015
7. AUTHOR(s) Y. Matsuyama, A. Buzo, and R. M. Gray		8. CONTRACT OR GRANT NUMBER(s) JSEP #N00014-75-C-0601 AFOSR #F44620-73-C-0065
9. PERFORMING ORGANIZATION NAME AND ADDRESS Stanford Electronics Laboratories Stanford University Stanford, Ca. 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Bldg. 410, Bolling AFB, DC 20332		12. REPORT DATE April 1978 13. NO. OF PAGES 54
14. MONITORING AGENCY NAME & ADDRESS (if diff. from Controlling Office) as above		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

speech compression, spectral distrotion measures, source coding, linear predictive coding

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In recent years several measures of distortion between speech waveforms have been proposed as substitutes for the traditional but subjectively inadequate mean-squared error. All of these measures involve some form of distortion measure between the second order properties of the speech processes producing the waveforms instead of an average of the waveform error power. In particular, they depend on the power spectral densities or linear models of the speech process. In this report the properties and interrelations of several such measures are developed. In particular, the relative strengths or equivalences of the various implications and

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

19. KEY WORDS (Continued)

20 ABSTRACT (Continued)

applications of those measures to prediction, detection, and coding are summarized.

DD FORM 1 JAN 73 1473 (BACK)

EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

JSEP REPORTS DISTRIBUTION LIST

Department of Defense

Director
National Security Agency
Attn: Dr. T. J. Beahn
Fort George G. Meade
Maryland 20755

Defense Documentation Center (12)
Attn: DDC-TCA (Mrs. V. Caponio)
Cameron Station
Alexandria, Virginia 22314

Assistant Director
Electronics and Computer Sciences
Office of Director of Defense
Research and Engineering
The Pentagon
Washington, D.C. 20315

Defense Advanced Research
Projects Agency
Attn: Dr. R. Reynolds
1400 Wilson Boulevard
Arlington, Virginia 22209

Department of the Army

Commandant
US Army Air Defense School
Attn: ATSAD-T-CSM
Fort Bliss, Texas 79916

Commander
US Army Armament R&D Command
Attn: DRDAR-TSS
Dover, New Jersey 07801

Commander
US Army Armament R&D Command (BRL)
Attn: DRDAR-TSB-S
Aberdeen Proving Ground
Aberdeen, Maryland 21005

Commandant
US Army Command and General Staff
College
Attn: Acquisitions, Library Division
Fort Leavenworth, Kansas 66027

Commander
US Army Communication Command
Attn: CC-OPS-PD
Fort Huachuca, Arizona 85613

Commander
US Army Materials and Mechanics
Research Center
Attn: Chief, Materials Science Div.
Watertown, Massachusetts 02172

Commander
US Army Materiel Development and
Readiness Command
Attn: Technical Library, Rm. 7S 35
5001 Eisenhower Avenue
Alexandria, Virginia 22333

Commander
US Army Missile R&D Command
Attn: Chief, Document Section
Redstone Arsenal, Alabama 35809

Commander
US Army Satellite Communications Agency
Fort Monmouth, New Jersey 07703

Director
US Army Signals Warfare Laboratory
Attn: DELSW-OS
Arlington Hall Station
Arlington, Virginia 22212

Project Manager
ARTADS
EAI Building
West Long Branch, New Jersey 07764

NOTE: One (1) copy to each addressee unless otherwise indicated.

Commander/Director
Atmospheric Sciences Lab. (ECOM)
Attn: DRSEL-BL-DD
White Sands Missile Range
New Mexico 88002

Commander
US Army Electronics Command
Attn: DRSEL-NL-O (Dr. H. S. Bennett)
Fort Monmouth, New Jersey 07703

Director
TRI-TAC
Attn: TT-AD (Mrs. Briller)
Fort Monmouth, New Jersey 07703

Commander
US Army Electronics Command
Attn: DRSEL-CT-L (Dr. R. Buser)
Fort Monmouth, New Jersey 07703

Director
Electronic Warfare Lab. (ECOM)
Attn: DRSEL-WL-MY
White Sands Missile Range
New Mexico 88002

Executive Secretary, TAC/JSEP
US Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

Commander
Harry Diamond Laboratories
Attn: Mr. John E. Rosenberg
2800 Powder Mill Road
Adelphi, Maryland 20783

HQDA (DAMA-ARZ-A)
Washington, D.C. 20310

Commander
US Army Electronics Command
Attn: DRSEL-TL-E (Dr. J. A. Kohn)
Fort Monmouth, New Jersey 07703

Commander
US Army Electronics Command
Attn: DRSEL-TL-EN
(Dr. S. Kroenenberg)
Fort Monmouth, New Jersey 07703

Commander
US Army Electronics Command
Attn: DRSEL-NL-T (Mr. R. Kulinyi)
Fort Monmouth, New Jersey 07703

Commander
US Army Electronics Command
Attn: DRSEL-NL-B (Dr. E. Lieblein)
Fort Monmouth, New Jersey 07703

Commander
US Army Electronics Command
Attn: DRSEL-TL-MM (Mr. N. Lipetz)
Fort Monmouth, New Jersey 07703

Commander
US Army Electronics Command
Attn: DRSEL-RD-O (Dr. W. S. McAfee)
Fort Monmouth, New Jersey 07703

Director
Night Vision Laboratory
Attn: DRSEL-NV-D
Fort Belvoir, Virginia 22060

Col. Robert Noce
Senior Standardization Representative
US Army Standardization Group, Canada
Canadian Force Headquarters
Ottawa, Ontario, CANADA KIA)K2

Commander
US Army Electronics Command
Attn: DRSEL-NL-B (Dr. D. C. Pearce)
Fort Monmouth, New Jersey 07703

Commander
US Army Electronics Command
Attn: DRSEL-NL-RH-1
(Dr. F. Schwering)
Fort Monmouth, New Jersey 07703

Commander
US Army Electronics Command
Attn: DRSEL-TL-I
(Dr. C. G. Thornton)
Fort Monmouth, New Jersey 07703

US Army Research Office (3)
Attn: Library
P. O. Box 12211
Research Triangle Park
North Carolina 27709

Director
Division of Neuropsychiatry
Walter Reed Army Institute
of Research
Washington, D.C. 20012

Commander
White Sands Missile Range
Attn: STEWS-ID-R
White Sands Missle Range
New Mexico 88002

Department of the Air Force

Mr. Robert Barrett
RADC/ETS
Hanscom AFB, Massachusetts 01731

Dr. Carl E. Baum
AFWL (ES)
Kirtland AFB, New Mexico 87117

Dr. E. Champagne
AFAL/DH
Wright-Patterson AFB, Ohio 45433

Dr. R. P. Dolan
RADC/ETSD
Hanscom AFB, Massachusetts 01731

Mr. W. Edwards
AFAL/TE
Wright-Patterson AFB, Ohio 45433

Professor R. E. Fontana
Head, Department of Electrical
Engineering
AFIT/ENE
Wright-Patterson AFB, Ohio 45433

Dr. Alan Garscadden
AFAPL/POD
Wright-Patterson AFB, Ohio 45433

USAF European Office of Aerospace
Research
Attn: Major J. Gorrell
Box 14, FPO, New York 09510

LTC Richard J. Gowen
Department of Electrical Engineering
USAF Academy, Colorado 80840

Mr. Murray Kesselman (ISCA)
Rome Air Development Center
Griffiss AFB, New York 13441

Dr. G. Knausenberger
Air Force Member, TAC
Air Force Office of Scientific
Research, (AFSC) AFSOR/NE
Bolling Air Force Base, DC 20332

Dr. L. Kravitz
Air Force Member, TAC
Air Force Office of Scientific
Research, (AFSC) AFSOR/NE
Bolling Air Force Base, DC 20332

Mr. R. D. Larson
AFAL/DHR
Wright-Patterson AFB, Ohio 45433

Dr. Richard B. Mack
RADC/ETER
Hanscom AFB, Massachusetts 01731

Mr. John Mottsmith (MCIT)
HQ ESD (AFSC)
Hanscom AFB, Massachusetts 01731

Dr. Richard Picard
RADC/ETSL
Hanscom AFB, Massachusetts 01731

Dr. J. Ryles
Chief Scientist
AFAL/CA
Wright-Patterson AFB, Ohio 45433

Dr. Allan Schell
RADC/ETE
Hanscom AFB, Massachusetts 01731

Mr. H. E. Webb, Jr. (ISCP)
Rome Air Development Center
Griffisss AFB, New York 13441

LTC G. Wepfer
Air Force Office of Scientific Research, (AFSC) AFOSR/NP
Bolling Air Force Base, DC 20332

LTC G. McKemie
Air Force Office of Scientific Research, (AFSC) AFOSR/NM
Bolling Air Force Base, DC 20332

Department of the Navy

Dr. R. S. Allgaier
Naval Surface Weapons Center
Code WR-303
White Oak
Silver Spring, Maryland 20910

Naval Weapons Center
Attn: Code 5515, H. F. Blazek
China Lake, California 93555

Dr. H. L. Blood
Technical Director
Naval Undersea Center
San Diego, California 95152

Naval Research Laboratory
Attn: Code 5200, A. Brodzinsky
4555 Overlook Avenue, SW
Washington, D.C. 20375

Naval Research Laboratory
Attn: Code 7701, J. D. Brown
4555 Overlook Avenue, SW
Washington, D.C. 20375

Naval Research Laboratory
Attn: Code 5210, J. E. Davey
4555 Overlook Avenue, SW
Washington, D.C. 20375

Naval Research Laboratory
Attn: Code 5460/5410, J. R. Davis
4555 Overlook Avenue, SW
Washington, D.C. 20375

Naval Ocean Systems Center
Attn: Code 75, W. J. Dejka
271 Catalina Boulevard
San Diego, California 92152

Naval Weapons Center
Attn: Code 601, F. C. Essig
China Lake, California 93555

Naval Research Laboratory
Attn: Code 5510, W. L. Faust
4555 Overlook Avenue, SW
Washington, D.C. 20375

Naval Research Laboratory
Attn: Code 2626, Mrs. D. Folen
4555 Overlook Avenue, SW
Washington, D.C. 20375

Dr. Robert R. Fossum
Dean of Research
Naval Postgraduate School
Monterey, California 93940

Dr. G. G. Gould
Technical Director
Naval Coastal System Laboratory
Panama City, Florida 32401

Naval Ocean Systems Center
Attn: Code 753, P. H. Johnson
271 Catalina Boulevard
San Diego, California 92152

Donald E. Kirk
Professor and Chairman
Electronic Engineer, SP-304
Naval Postgraduate School
Monterey, California 93940

Naval Air Development Center
Attn: Code 01, Dr. R. K. Lobb
Johnsville
Warminster, Pennsylvania 18974

Naval Research Laboratory
Attn: Code 5270, B. D. McCombe
4555 Overlook Avenue, SW
Washington, D.C. 20375

Capt. R. B. Meeks
Naval Sea Systems Command, NC #3
2531 Jefferson Davis Highway
Arlington, Virginia 20362

Dr. H. J. Mueller
Naval Air Systems Command
Code 310, JP #1
1411 Jefferson Davis Highway
Arlington, Virginia 20360

Dr. J. H. Mills, Jr.
Naval Surface Weapons Center
Electronics Systems Department
Code DF
Dahlgren, Virginia 22448

Naval Ocean Systems Center
Attn: Code 702, H. T. Mortimer
271 Catalina Boulevard
San Diego, California 92152

Naval Air Development Center
Attn: Technical Library
Johnsville
Warminster, Pennsylvania 18974

Naval Ocean Systems Center
Attn: Technical Library
271 Catalina Boulevard
San Diego, California 92152

Naval Research Laboratory
Underwater Sound Reference Division
Technical Library
P. O. Box 8337
Orlando, Florida 32806

Naval Surface Weapons Center
Attn: Technical Library
Code DX-21
Dahlgren, Virginia 22448

Naval Surface Weapons Center
Attn: Technical Library
Building 1-330, Code WX-40
White Oak Laboratory
Silver Spring, Maryland 20910

Naval Training Equipment Center
Attn: Technical Library
Orlando, Florida 32813

Naval Undersea Center
Attn: Technical Library
San Diego, California 92152

Naval Underwater Systems Center
Attn: Technical Library
Newport, Rhode Island 02840

Office of Naval Research
Electronic and Solid State
Sciences Program (Code 427)
800 North Quincy Street
Arlington, Virginia 22217

Office of Naval Research
Mathematics Program (Code 432)
800 North Quincy Street
Arlington, Virginia 22217

Office of Naval Research
Naval Systems Division
Code 220/221
800 North Quincy Street
Arlington, Virginia 22217

Director
Office of Naval Research
New York Area Office
715 Broadway, 5th Floor
New York, New York 10003

Office of Naval Research
San Francisco Area Office
One Hallidie Plaza, Suite 601
San Francisco, California 94102

Director
Office of Naval Research Branch Office
495 Summer Street
Boston, Massachusetts 02210

Director
Office of Naval Research Branch Office
536 South Clark Street
Chicago, Illinois 60605

Director
Office of Naval Research Branch Office
1030 East Green Street
Pasadena, California 91101

Mr. H. R. Riedl
Naval Surface Weapons Center
Code WR-34
White Oak Laboratory
Silver Spring, Maryland 20910

Naval Air Development Center
Attn: Code 202, T. J. Shopple
Johnsville
Warminster, Pennsylvania 18974

Naval Research Laboratory
Attn: Code 5403, J. E. Shore
4555 Overlook Avenue, SW
Washington, D.C. 20375

A. L. Slafkovsky
Scientific Advisor
Headquarters Marine Corps
MC-RD-1, Arlington Annex
Washington, D.C. 20380

Harris B. Stone
Office of Research, Development,
Test and Evaluation
NOP-987
The Pentagon, Room 5D760
Washington, D.C. 20350

Mr. L. Sumney
Naval Electronics Systems Command
Code 3042, NC #1
2511 Jefferson Davis Highway
Arlington, Virginia 20360

David W. Taylor
Naval Ship Research and
Development Center
Code 522.1
Bethesda, Maryland 20084

Naval Research Laboratory
Attn: Code 4105, Dr. S. Teitler
4555 Overlook Avenue, SW
Washington, D.C. 20375

Lt. Cdr. John Turner
NAVMAT 0343
CP #5, Room 1044
2211 Jefferson Davis Highway
Arlington, Virginia 20360

Naval Ocean Systems Center
Attn: Code 746, H. H. Wieder
271 Catalina Boulevard
San Diego, California 92152

Dr. W. A. Von Winkle
Associate Technical Director for
Technology
Naval Underwater Systems Center
New London, Connecticut 06320

Dr. Gernot M. R. Winkler
Director, Time Service
US Naval Observatory
Massachusetts Ave. at 34th St., NW
Washington, D.C. 20390

Other Government Agencies

Dr. Howard W. Etzel
Deputy Director
Division of Materials Research
National Science Foundation
1800 G Street
Washington, D.C. 20550

Mr. J. C. French
National Bureau of Standards
Electronics Technology Division
Washington, D.C. 20234

Dr. Jay Harris
Program Director
Devices and Waves Program
National Science Foundation
1800 G Street
Washington, D.C. 20550

Los Alamos Scientific Laboratory
Attn: Reports Library
P. O. Box 1663
Los Alamos, New Mexico 87544

Dr. Dean Mitchell
Program Director
Solid-State Physics
Division of Materials Research
National Science Foundation
1800 G Street
Washington, D.C. 20550

Mr. F. C. Schwenk, RD-T
National Aeronautics and Space
Administration
Washington, D.C. 20546

M. Zane Thornton
Deputy Director, Institute for
Computer Sciences and Technology
National Bureau of Standards
Washington, D.C. 20234

Director
Stanford Electronics Laboratories
Stanford University
Stanford, California 94305

Nongovernment Agencies

Director
Columbia Radiation Laboratory
Columbia University
538 West 120th Street
New York, New York 10027

Director
Coordinated Science Laboratory
University of Illinois
Urbana, Illinois 61801

Director of Laboratories
Division of Engineering and
Applied Physics
Harvard University
Pierce Hall
Cambridge, Massachusetts 02138

Director
Electronics Research Center
The University of Texas
Engineering-Science Bldg. 112
Austin, Texas 78712

Director
Electronics Research Laboratory
University of California
Berkeley, California 94720

Director
Electronics Sciences Laboratory
University of Southern California
Los Angeles, California 90007

Director
Microwave Research Institute
Polytechnic Institute of New York
333 Jay Street
Brooklyn, New York 11201

Director
Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Officer in Charge
Carderock Laboratory
Code 18, G. H. Gleissner
David Taylor Naval Ship Research
and Development Center
Bethesda, Maryland 20084

Dr. Roy F. Potter
3868 Talbot Street
San Diego, California 92106